

## LJMU Research Online

**Marasco, A, Crain, RA, Schaye, J, Bahe, YM, van der Hulst, T, Theuns, T and Bower, RG**

**The environmental dependence of H I in galaxies in the EAGLE simulations**

<http://researchonline.ljmu.ac.uk/id/eprint/4664/>

### Article

**Citation** (please note it is advisable to refer to the publisher's version if you intend to cite from this work)

**Marasco, A, Crain, RA, Schaye, J, Bahe, YM, van der Hulst, T, Theuns, T and Bower, RG (2016) The environmental dependence of H I in galaxies in the EAGLE simulations. *Monthly Notices of the Royal Astronomical Society*, 461 (3). pp. 2630-2649. ISSN 0035-8711**

LJMU has developed [LJMU Research Online](http://researchonline.ljmu.ac.uk/) for users to access the research output of the University more effectively. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LJMU Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain.

The version presented here may differ from the published version or from the version of the record. Please see the repository URL above for details on accessing the published version and note that access may require a subscription.

For more information please contact [researchonline@ljmu.ac.uk](mailto:researchonline@ljmu.ac.uk)

# The environmental dependence of H I in galaxies in the EAGLE simulations

Antonino Marasco,<sup>1</sup>★ Robert A. Crain,<sup>2</sup> Joop Schaye,<sup>3</sup> Yannick M. Bahé,<sup>4</sup>  
Thijs van der Hulst,<sup>1</sup> Tom Theuns<sup>5</sup> and Richard G. Bower<sup>5</sup>

<sup>1</sup>Kapteyn Astronomical Institute, Postbus 800, NL-9700 AV Groningen, the Netherlands

<sup>2</sup>Astrophysics Research Institute, Liverpool John Moores University, 146 Brownlow Hill, Liverpool L3 5RF, UK

<sup>3</sup>Leiden Observatory, Leiden University, PO Box 9513, NL-2300 RA Leiden, the Netherlands

<sup>4</sup>Max-Planck-Institut für Astrophysik, Karl-Schwarzschild Str. 1, D-85748 Garching, Germany

<sup>5</sup>Institute for Computational Cosmology, Department of Physics, University of Durham, South Road, Durham DH1 3LE, UK

Accepted 2016 June 18. Received 2016 June 10; in original form 2016 March 22

## ABSTRACT

We use the EAGLE suite of cosmological hydrodynamical simulations to study how the H I content of present-day galaxies depends on their environment. We show that EAGLE reproduces observed H I mass–environment trends very well, while semi-analytic models typically overpredict the average H I masses in dense environments. The environmental processes act primarily as an on/off switch for the H I content of satellites with  $M_* > 10^9 M_\odot$ . At a fixed  $M_*$ , the fraction of H I-depleted satellites increase with increasing host halo mass  $M_{200}$  in response to stronger environmental effects, while at a fixed  $M_{200}$  it decreases with increasing satellite  $M_*$  as the gas is confined by deeper gravitational potentials. H I-depleted satellites reside mostly, but not exclusively, within the virial radius  $r_{200}$  of their host halo. We investigate the origin of these trends by focusing on three environmental mechanisms: ram pressure stripping by the intragroup medium, tidal stripping by the host halo and satellite–satellite encounters. By tracking back in time the evolution of the H I-depleted satellites, we find that the most common cause of H I removal is satellite encounters. The time-scale for H I removal is typically less than 0.5 Gyr. Tidal stripping occurs in haloes of  $M_{200} < 10^{14} M_\odot$  within  $0.5 \times r_{200}$ , while the other processes act also in more massive haloes, generally within  $r_{200}$ . Conversely, we find that ram pressure stripping is the most common mechanism that disturbs the H I morphology of galaxies at redshift  $z = 0$ . This implies that H I removal due to satellite–satellite interactions occurs on shorter time-scales than the other processes.

**Key words:** methods: numerical – galaxies: clusters: general – galaxies: evolution – galaxies: interactions – galaxies: ISM.

## 1 INTRODUCTION

It has long been recognized that the properties of galaxies depend significantly on their environment. Optical studies were the first to shed light on the effect of the local density on the properties of galaxies. Oemler (1974) and Dressler (1980) first established the existence of a clear morphology–density relation, indicating a steady increase in the population of elliptical/S0 systems, and a corresponding decrease of spiral galaxies, in environments of increasing density. Later, with the advent of a large photometric and spectroscopic data base provided by the Sloan Digital Sky Survey (SDSS; York et al. 2000), a systematic study of the environmental effect on the colour, star formation and the structure of galaxies, became possible for a variety of densities and cosmic epochs (e.g.

Balogh et al. 2004; Cooper et al. 2006; Wetzel, Tinker & Conroy 2012). In particular, it was highlighted that the specific star formation rate (sSFR) of galaxies depends on their environment, with an average decrease of one order of magnitude in sSFR when moving from more isolated systems to the densest regions (Kauffmann et al. 2004) at a given stellar mass. Simultaneously, different studies emphasized how galaxy properties depend primarily on the stellar mass of the system: lower mass galaxies preferentially exhibit discy morphologies with a young (blue) stellar population, while massive systems are mainly spheroidal with old (red) stars (Kauffmann et al. 2003; Baldry et al. 2004). These findings highlight the importance of disentangling the role of ‘nature’, i.e. internal processes, and ‘nurture’, i.e. the ensemble of environmental processes.

It is reasonable to assume that, whichever are the environmental processes that affect the star formation in a galaxy, they must also – and presumably first – affect the gas (neutral hydrogen) content of the system, which is the material from which stars are formed.

★ E-mail: marasco@astro.rug.nl

In particular, atomic hydrogen (H I) is known to be particularly sensitive to the environment: H I discs are often much more extended than their optical counterparts and are therefore more sensitive to external influences (e.g. Yun, Ho & Lo 1994).

Early H I environmental studies focused on galaxies in clusters, and revealed that these systems are significantly more H I deficient than those in the field, and that the magnitude of the H I deficiency correlates with the distance from the cluster centre (Giovanelli & Haynes 1985; Solanes et al. 2001). Resolved H I observations of the Virgo Cluster (e.g. Chung et al. 2009) showed the presence of small, truncated H I discs within 0.5 Mpc from the cluster core and head-tail H I morphologies for galaxies at larger ( $\sim 1$  Mpc) distances. It became clear that ram pressure stripping by the intracluster medium (Gunn & Gott 1972) is an important mechanism of gas removal (e.g. Oosterloo & van Gorkom 2005; Jaffé et al. 2015).

The advent of blind H I surveys such as the Arecibo Legacy Fast ALFA (ALFALFA; Giovanelli et al. 2005) enabled to extend the H I environmental studies of to a wider range of environmental densities. Fabello et al. (2012, hereafter Fab12) used stacked ALFALFA spectra extracted from optically selected galaxies and found that the average H I gas fraction in galaxies decreases faster than the sSFR with increasing environmental density, consistent with a scenario in which the environment affects first the (extended) H I disc and only later the (more concentrated) star-forming gas reservoir. From the comparison with the predictions of semi-analytical (SA) models of galaxy evolution, Fab12 concluded that ram pressure stripping of H I discs is already effective in group environments with halo masses larger than  $\sim 10^{13} M_{\odot}$ . The same conclusion was found by Catinella et al. (2013, hereafter Cat13) by cross-matching the GALEX Arecibo SDSS (GASS; Catinella et al. 2010) sample with the SDSS Data Release 7 (DR7) group catalogue of Yang et al. (2012). These findings favour a scenario where the H I content of galaxies varies smoothly across all environmental densities, with the classical field-cluster dichotomy representing only the two extremes of a more continuous trend.

Given that the environment acts on galaxies via hydrodynamical and gravitational forces, numerical simulations of galaxy evolution constitute a powerful tool to tackle its study. Although several simulations have followed the evolution of galaxies in a single environment (e.g. Tonnesen, Bryan & van Gorkom 2007; Limousin et al. 2009; Few et al. 2012; Villalobos et al. 2012), for a systematic study of the effect of different densities on galaxies of different masses, large-scale cosmological simulations are needed. Dark-matter-only cosmological simulations and SA models self-consistently follow only the tidal forces that dark matter structures experience and cannot provide a complete view of the role of the environment. The advantage of cosmological hydrodynamical simulations over SA models is that they self-consistently follow both the gravitational and the hydrodynamical processes, provided they have sufficient resolution and dynamical range to make predictions for the interplay between galaxies and their environment.

In this paper we make use of the ‘Evolution and Assembly of Galaxies and their Environments’ (EAGLE; Crain et al. 2015; Schaye et al. 2015), a suite of state-of-the-art cosmological hydrodynamical simulations, to study how the environment impacts the H I content and morphology of galaxies in the  $z = 0$  Universe. EAGLE has been shown to reproduce several observed scaling relations of galaxies, such as the Tully–Fisher and the mass–star formation rate relations (Schaye et al. 2015), and predicts present-day galaxy colours (Trayford et al. 2015) and stellar mass assembly histories (Furlong et al. 2015) that agree with the observations very well. The properties of neutral gas in the EAGLE simulations have been explored in a

number of works: Rahmati et al. (2016) studied the H I column density distribution and covering fraction in the circumgalactic medium around high-redshift galaxies; Bahé et al. (2016) focused on the H I morphology and distribution in present-day galaxies; Crain et al. (2016) explored the impact of mass resolution and feedback/star formation efficiency on the overall properties of galactic H I; Lagos et al. (2015, 2016) analysed the galactic molecular hydrogen and its connection to ‘the Fundamental Plane of star formation’.

For our purposes, one of the strengths of the EAGLE simulations is that the ill-constrained efficiency of feedback processes have been calibrated against the stellar properties of the overall galaxy population, and neither gas properties nor environmental trends were considered during the calibration. The environmental effects that we aim to study are thus governed by gravitational and hydrodynamical forces that are modelled explicitly and followed self-consistently. Our approach to this study is analogous to that adopted in other works that made use of this suite of simulations: we first verify whether or not EAGLE predicts the observed scaling relations with the environment, and then we use its predictive power both to extend these relations to regimes that are not yet accessible observationally and to investigate their origin.

The paper is structured as follows. In Section 2 we present a brief description of the simulations and describe the method adopted to derive the H I content of simulated galaxies. In Section 3 we compare the predictions of EAGLE and of three SA models of galaxy evolution with the observations of Fab12 and Cat13, showing that EAGLE is in much better agreement with the observations than the SA models. In Section 4 we extend the predicted trends to a larger dynamical range of stellar masses and environment densities than has been observed. We discuss our results in Section 5, where we investigate the mechanisms by which the environment influences the H I of galaxies. Finally, our conclusions are presented in Section 6.

## 2 SIMULATIONS

A detailed description of the EAGLE simulations is presented by Schaye et al. (2015) and Crain et al. (2015). Here, we briefly summarize their main characteristics. EAGLE is a suite of cosmological hydrodynamical simulations performed in a standard  $\Lambda$  cold dark matter ( $\Lambda$ CDM) framework, adopting the cosmological parameters inferred from the first-year *Planck* data (Planck Collaboration XVI 2014). The run that we consider for most of this paper is the largest simulation available in EAGLE (Ref-L100N1504). It follows the evolution of  $1504^3$  dark matter and gas particles in a cubic box of side length 100 comoving Mpc (cMpc) from redshift  $z = 127$  to the present time via a modified version of the  $N$ -body+smoothed particle hydrodynamics (SPH) code GADGET-3 (Springel 2005). The particle mass is  $9.7 \times 10^6 M_{\odot}$  for the dark matter and  $1.81 \times 10^6 M_{\odot}$  (initially) for the baryonic component. The gravitational softening length is 0.7 proper kpc below redshift  $z = 2.8$ .

EAGLE uses a formulation of the SPH, known as ANARCHY (Dalla Vecchia, in preparation, see also appendix A of Schaye et al. 2015 and Schaller et al. 2015), which alleviates significantly the issues related to artificial gas clumping and the poor treatment of hydrodynamical instabilities associated with the classical SPH scheme (discussed by e.g. Kaufmann et al. 2006; Agertz et al. 2007). ANARCHY also uses the artificial viscosity switch from Cullen & Dehnen (2010), an artificial conduction switch analogue to that of Price (2008) and the time step limiter proposed by Durier & Dalla Vecchia (2012). ANARCHY is similar to the SPH implementation SPH-Gal (Hu et al. 2014), for which hydrodynamical tests such as the Kelvin–Helmholtz instability and the blob test produce outcomes

comparable to those of grid hydrodynamical codes. We acknowledge that the blob test presented by Hu et al. (2014) is truly representative neither for the resolution of the *EAGLE* Ref-L100N1504 run, which is typically two orders of magnitude lower, nor for the typical density contrast between the interstellar medium (ISM) and the intergalactic gas. On the other hand, the outcome of an environmental process like ram pressure stripping is driven by the competition between ram pressure and gravity, and a blob test is only partially relevant to it as it does not incorporate the latter.

*EAGLE* incorporates recipes for subgrid physics, including the star formation implementation of Schaye & Dalla Vecchia (2008), star formation feedback in thermal form based on the prescription of Dalla Vecchia & Schaye (2012), radiative gas cooling and photo-heating for 11 different elements from Wiersma, Schaye & Smith (2009a), stellar mass loss from Wiersma et al. (2009b) and accreting supermassive black holes and active galactic nuclei (AGN) feedback in thermal form based on Springel, Di Matteo & Hernquist (2005b), Booth & Schaye (2009) and Rosas-Guevara et al. (2015). Given that *EAGLE* does not model the cold gas phase, a global temperature floor  $T_{\text{eos}}$  for gas particles is imposed, corresponding to the equation of state  $P \propto \rho^{(4/3)}$ , normalized to 8000 K at a density of  $n_{\text{H}} = 0.1 \text{ cm}^{-3}$ . Gas particles are eligible to form star particles when they have cooled enough to reach temperatures  $\log_{10}(T) < \log_{10}(T_{\text{eos}}) + 0.5$ , and densities  $n_{\text{H}} > n_{\text{H}}^*(Z)$ , where  $n_{\text{H}}^*(Z)$  is a threshold that depends on the metallicity as described in Schaye et al. (2015).

The free parameters associated with feedback are calibrated to reproduce three key observables at redshift zero: the galaxy stellar mass function, the size–mass relation of disc galaxies and the galaxy–black hole mass relation, which match the observations with an accuracy that is unprecedented for hydrodynamical simulations. Details of the calibration procedure are presented by Crain et al. (2015). The galaxy stellar mass function for the *EAGLE* run Ref-L100N1504 is in excellent agreement with the observations for stellar masses between  $10^9$  and  $10^{11.5} M_{\odot}$ . In our analysis, we will focus on systems in this range of stellar masses.

## 2.1 Computing the atomic hydrogen masses

Evaluating the  $\text{H I}$  masses of galaxies in *EAGLE* is a non-trivial task, as the simulation was neither designed to account for the effect of self-shielding on the neutral/ionized phases of hydrogen nor to keep track of its atomic/molecular content. For consistency with previous works that used the same suite of simulations, we decided to follow the  $\text{H I}$  prescription adopted by Bahé et al. (2016) and Crain et al. (2016), which we summarize in the following.

For each gas particle in the simulated box, we compute the fraction of hydrogen that is neutral ( $\text{H I} + \text{H}_2$ ) by using the redshift-dependent fitting formula of Rahmati et al. (2013, see their table A1), which is calibrated using smaller simulations with detailed radiation transport modelling performed via *TRAPHIC* (Pawlik & Schaye 2008) and consider gas to be in (photo+collisional) ionization equilibrium with ultraviolet (UV) background with  $\text{H I}$  photoionization rate of  $\Gamma_{\text{H I}} = 8.34 \times 10^{-14} \text{ s}^{-1}$  at  $z = 0$  (Haardt & Madau 2001). As shown by Crain et al. (2016), variation of  $\Gamma_{\text{H I}}$  by a factor of a few has virtually no impact on the neutral gas content of simulated galaxies at  $z = 0$ . The collisional ionization rate of star-forming particles is computed by fixing their temperature to a value of  $10^4 \text{ K}$ , characteristic of the warm phase of the ISM, rather than using their SPH temperature, which only reflects

the effective pressure of a multiphase ISM.<sup>1</sup> In Ref-L100N1504, approximately 50 per cent of the neutral gas mass associated with present-day galaxies with  $M_* > 10^9 M_{\odot}$  is contributed by star-forming particles, thus such a temperature correction is relevant to our study.

If a gas particle is eligible for star formation, we partition the fraction of neutral hydrogen into atomic ( $\text{H I}$ ) and molecular ( $\text{H}_2$ ) forms as  $\frac{n_{\text{H}_2}}{n_{\text{H I}}} \simeq (\frac{P}{P_0})^{\alpha}$  (Blitz & Rosolowsky 2006), where  $n$  is the gas volume density,  $P$  is the gas pressure,  $P_0 = 4.3 \times 10^4 \text{ cm}^{-3} \text{ K}$  and  $\alpha = 0.92$ . Note that this partitioning is not unique, different approaches are indeed possible (e.g. Lagos et al. 2015) and can lead to somewhat different results. The procedure adopted here results in galactic  $\text{H I}$  discs whose mass and size are in good agreement with the observations for  $M_* > 10^{10} M_{\odot}$  (Bahé et al. 2016), but slightly too  $\text{H I}$  deficient in less massive galaxies (Crain et al. 2016). We do not model directly the potentially significant influence of local radiation sources on the neutral fraction, but we point out that their impact on  $\text{H}_2$  fractions is implicitly accounted for by the empirical Blitz & Rosolowsky (2006) relation.

In this work, the total  $\text{H I}$  mass of each simulated galaxy – defined as a self-bound subhalo identified via the *SUBFIND* algorithm (Dolag et al. 2009) – is determined in two ways. The first is to sum the  $\text{H I}$  content of all gas particles that belong to that particular subhalo. This method gives the total  $\text{H I}$  content of each subhalo and is preferred when we are interested in what the simulation predicts about the gas bound to a particular system. We use this method in Sections 4 and 5.  $\text{H I}$  observations, however, know little about whether some gas is bound to a galaxy or not. Therefore, the second method is to sum the  $\text{H I}$  content of all particles within a given 2D circular aperture and a given line-of-sight velocity range from the centre of potential of each subhalo. This provides a ‘biased’ estimate of the  $\text{H I}$  content of a galaxy which is directly comparable to  $\text{H I}$  surveys like ALFALFA. We adopt this method in Section 3.

## 3 COMPARISON WITH OBSERVATIONS

In this section, we compare the observations of *Fab12* and *Cat13* with the predictions of *EAGLE*. It is important to reiterate that the *EAGLE* simulations have not been calibrated to reproduce the observed properties of the ISM of galaxies. Therefore, the trends that we report can be considered predictions of the simulation.

We include in our comparison the predictions of three current SA models of galaxy formation: those of Guo et al. (2011), Guo et al. (2013) and Henriques et al. (2015), hereafter *Guo11*, *Guo13* and *Hen15*, respectively. These models are based on the Millennium (Springel et al. 2005a) and Millennium-II (Boylan-Kolchin et al. 2009) cosmological simulations, which assume the 1-year *Wilkinson Microwave Anisotropy Probe* (*WMAP1*) cosmology (Spergel et al. 2003), and they each incorporate the treatment of environmental processes via a number of analytic recipes. In these models, subhaloes have a reservoir of hot gas at the virial temperature that steadily cools on to the galactic disc. The details of this process depend on the halo mass and on the metallicity of the hot gas. When

<sup>1</sup> In Crain et al. (2016), a temperature of  $10^4 \text{ K}$  is assigned also to particles that are not star forming, but whose pressure is still within 0.5 dex from that imposed by the equation of state and for which  $T_{\text{eos}} > 10^4 \text{ K}$ . This is to avoid high density, low metallicity gas – which may be not star forming – from having an unrealistically high ionization fraction. In this study we do not make use of such refinement. However, we verified that the impact of this correction on the  $\text{H I}$  content of galaxies is negligible.



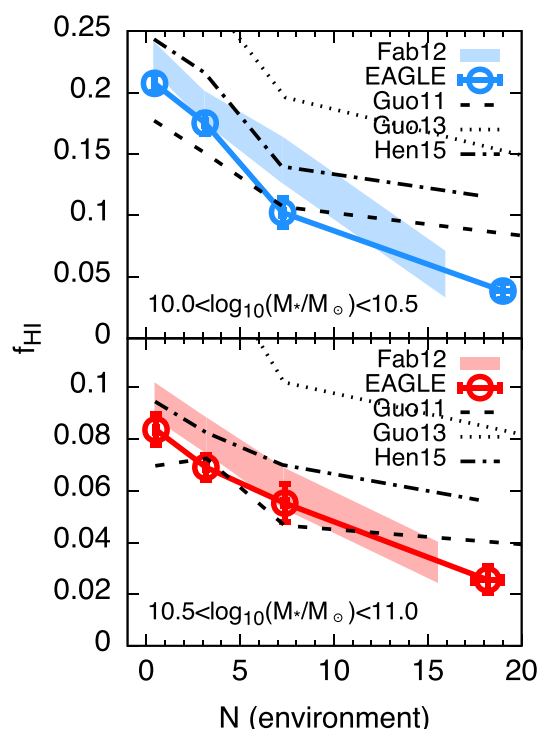
a subhalo crosses the virial radius of a neighbouring friends-of-friends (FoF) group, environmental processes are switched on: tidal forces and ram pressure stripping remove material from the subhalo's gas reservoir and deposit it on to that of the central galaxy of the group. The infalling subhalo is then left with less gas (or no gas at all) available for cooling and fuelling future star formation. It is important to note that the disc of cold gas is assumed to remain unaffected by this process, unless the subhalo is physically disrupted by tidal interactions. This assumption is the key to interpret the differences that we will show between EAGLE and the SA models. The models of galaxy formation implemented by Guo11 and Guo13 are virtually the same, but in the latter the underlying dark matter simulation has been re-scaled to the WMAP7 cosmology (Komatsu et al. 2011), and the parameters of the model have been re-calibrated to reproduce the same diagnostics considered by Guo11. In Hen15 the underlying simulation is re-scaled instead to the first-year Planck cosmology (Planck Collaboration XVI 2014), details of the star formation threshold and the reincorporation of gas ejected by galactic winds have been modified with respect to Guo13, and ram pressure stripping is artificially suppressed in groups with virial mass below  $10^{14} M_{\odot}$ .

We estimate the H I masses from the ‘cold gas’ masses predicted by the SA models by using one of the prescriptions described by Lagos et al. (2011, see their section 2.3), specifically that based on the star formation law of Blitz & Rosolowsky (2006). This recipe uses the stellar and the cold gas surface density profiles to compute the H I and the H<sub>2</sub> surface density profiles, which we derive in each galaxy from its centre to five times the gas scale length, and then integrate to determine the total masses. The recipe gives H<sub>2</sub>-to-H I ratios as a function of the stellar mass that are in good agreement with the observations (Lagos et al. 2011) and affords a more straightforward comparison with EAGLE, since the prescriptions for the molecular fraction are based on the same law. Note that, contrary to SA models, EAGLE does not ignore the presence of cold ( $T \sim 10^4$  K) ionized gas. As a consistency check, we also computed the atomic and molecular gas fractions by using a simple recipe based on the classical Kennicutt (1998) relation, which we inverted to derive the molecular gas masses. This yields H I masses that are about 50 per cent larger from those computed with the Lagos et al. recipe, but has overall little impact on the results presented here.

### 3.1 Comparison with stacked ALFALFA observations

Fab12 used stacked H I spectra extracted from the ALFALFA 40 per cent data set (Haynes et al. 2011) to infer the average H I mass fraction ( $M_{\text{H I}}/M_*$ , or  $f_{\text{H I}}$ ) and the sSFR as a function of the local galaxy density. The galaxies considered were selected from the GASS (Catinella et al. 2010). Stellar masses and star formation rates (SFRs) are derived from spectral energy distribution (SED) fitting using a Chabrier (2003) initial mass function (IMF). The environment density estimator used by Fab12 is the number  $N$  of galaxies with stellar mass above  $10^{9.5} M_{\odot}$  within a projected radius of 1 Mpc and within  $\pm 500 \text{ km s}^{-1}$  in line-of-sight velocity from a given object. The analysis is performed separately for two different bins in stellar masses,  $10^{10} < M_*/M_{\odot} < 10^{10.5}$  and  $10^{10.5} < M_*/M_{\odot} < 10^{11}$ , in order to break the mass–environment degeneracy.

H I masses in the simulation are computed by considering circular apertures of 150 kpc in diameter, corresponding to the Arecibo beam size at the median redshift of the observed sample (Giovanelli et al. 2005; Catinella et al. 2010), and line-of-sight velocity ranges



**Figure 1.**  $f_{\text{H I}} \equiv M_{\text{H I}}/M_*$  as a function of the environment parameter  $N$  (see text) for galaxies in the stellar mass range  $10 < \log_{10}(M_*/M_{\odot}) < 10.5$  (top panel) and  $10.5 < \log_{10}(M_*/M_{\odot}) < 11$  (bottom panel). Shaded regions represent data from Fab12, symbols connected with solid lines show the prediction of EAGLE, the other lines show the SA models of Guo11, Guo13 and Hen15. Error bars and the thickness of the shaded regions represent the  $1\sigma$  uncertainty on the mean and are derived by bootstrap resampling the galaxies in each bin. Note the different y-axis ranges. In all cases,  $f_{\text{H I}}$  decreases with  $N$  at given  $M_*$ . EAGLE is in better agreement with the observations than the SA models are.

of  $\pm 400 \text{ km s}^{-1}$  by analogy with the observations (see Section 2.1). The stacking technique adopted by Fab12 combines data for galaxies at different redshifts with different signal-to-noise ratio and may potentially introduce a bias in the calculation of the H I fractions which is difficult to mimic precisely in the simulations. Therefore, we do not attempt to implement an analogous stacking technique in the simulation. Stellar masses are computed by using a spherical aperture of radius 30 kpc. Galaxy pairs separated by less than a half-mass radius of the larger system are considered as single systems (see Schaye et al. 2015). Finally, the environment density estimator  $N$  is computed as in the observations: projected distances are evaluated in the  $(x, y)$  plane and  $v_z$  gives the line-of-sight velocities.

Fig. 1 shows the mean  $f_{\text{H I}}$  as a function of the environment density  $N$  in EAGLE (circles+solid line) and SA models (dashed lines) for two different bins in stellar mass, and compare it to the data of Fab12 (shaded region, Catinella, private communication). In all cases, four bins of  $N$  are considered:  $0 \leq N \leq 1$ ,  $2 \leq N \leq 5$ ,  $6 \leq N \leq 9$  and  $N \geq 10$ . Error bars and the thickness of the shaded region represent the  $1\sigma$  uncertainty on the mean at a given bin of  $N$  and are computed by bootstrap resampling the systems in each bin of  $N$  for both observations and simulations. In Fab12, H I fractions were normalized to the peak value to focus primarily on the H I trend with the environment. Here, we prefer to show the unnormalized values in order to demonstrate the similarity of the simulations to the observations. Note that the vertical scale is different in the two

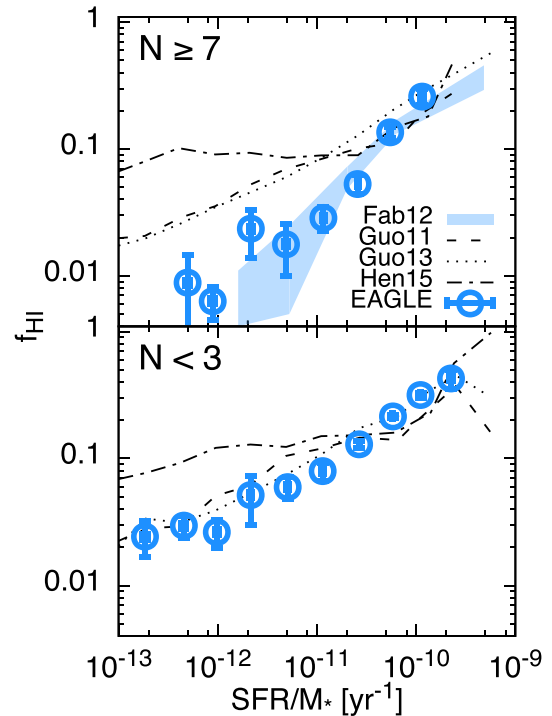
panels: on average, the  $H\text{I}$  fraction in galaxies in the more massive bin is about half of that in the lower mass bin.

From Fig. 1 it seems that both *EAGLE* and the SA models of *Guo11* and *Hen15* predict values of  $f_{H\text{I}}$  that are in overall agreement with the observations. *EAGLE* and *Hen15* slightly underpredict the  $H\text{I}$  fraction in all galaxies with  $N < 10$  (see also Fig. 3), whereas *Guo11* overpredict it at larger  $N$ . An exception is the SA model of *Guo13*, for which the  $H\text{I}$  fractions are almost a factor of 2 greater than observed. The reason for this behaviour is unclear. However, when we focus on the trend with the environment, *EAGLE* is in better agreement with the observations than the SA models. In fact, for the low stellar mass bin, both observations and *EAGLE* indicate that  $f_{H\text{I}}$  in galaxies drops by a factor of 4.5 when moving from isolated systems ( $N = 0$ ) to high-density environments ( $N \gtrsim 15$ ), whereas the drop predicted by SA models is only a factor of 2–2.5. Similarly, for the high stellar mass bin, the observations and *EAGLE* show a drop of a factor of 3, compared to a factor of only 1.5–2.0 in the SA models. The magnitude of this drop is underestimated not only in SA models, but also in other hydrodynamical cosmological runs like those of Rafieferantsoa et al. (2015, although we point out that they use a different post-processing scheme to calculate  $H\text{I}$ ).

Note that the rightmost points of each panel in Fig. 1 are displaced towards larger  $N$  with respect to the shaded area. In the SA models of *Guo11* and *Guo13*, in particular, the rightmost point exceeds the plot boundary and it is not shown. Both in the simulations and in the data, this point refers to the average  $N$  for systems with  $N \geq 10$ , thus suggesting that the run Ref-L100N1504 overpredicts the galaxy clustering in the high-density regimes with respect to *Fab12* observed sample. The presence of intrinsic differences in the environmental distribution of our sample with respect to that of *Fab12* can introduce a bias in our analysis. Specifically, by inspecting the distribution of  $N$ , we find that *EAGLE* predicts an excess of systems at  $N \geq 5$  and a deficit at lower  $N$  with respect to the observations. The SA models show a similar behaviour to *EAGLE*. An investigation of the origin of this discrepancy is beyond the scope of this study, so we attempt only to correct for this bias by defining an ‘unbiased’ subsample of *EAGLE* galaxies that follows the same distribution of  $N$  as the data. This subsample is built by using an iterative procedure where, at each step, we randomly extract an *EAGLE* galaxy from the bin of  $N$  with the largest (positive) difference between the fraction of systems in the simulation and in the data. This system is removed from the sample, the distribution of  $N$  is re-evaluated and extractions continue until the differences between the observed and simulated distributions are minimized. The trend of  $f_{H\text{I}}$  with  $N$  for the resulting unbiased subsample is fairly consistent with that shown in Fig. 1 for the full sample, with the obvious difference that now – by construction – the average  $N$  predicted and observed coincide.

Fig. 2 shows the trend between  $f_{H\text{I}}$  and sSFR for galaxies with  $10 < \log_{10}(M_*/M_\odot) < 11$  located in dense ( $N \geq 7$ , top panel) and sparse ( $N < 3$ , bottom panel) environments. The comparison with the observed trend is only available for the high-density environment. Here, *EAGLE* agrees remarkably well with the observed trend. SA models fail in the region where the sSFR drops below  $5 \times 10^{-11} \text{ yr}^{-1}$ , markedly overpredicting the mean  $H\text{I}$  fractions for galaxies in this regime. For comparison, galaxies with  $N < 3$  seem to settle at a higher  $f_{H\text{I}}$  with respect to the high-density systems with the same sSFR. Here, the difference between *EAGLE* and the SA models is smaller.

The different predictions of *EAGLE* in the two environments can be due to a number of things. First, it is possible that galaxies that live in denser environments have, on average, a greater  $M_*$

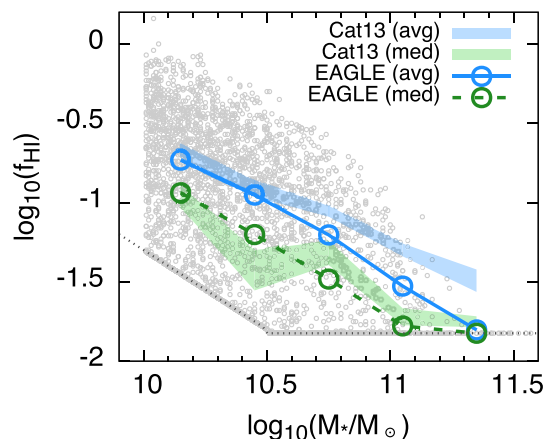


**Figure 2.**  $f_{H\text{I}}$  as a function of the sSFR for galaxies in environment densities  $N \geq 7$  (top panel) and  $N < 3$  (bottom panel). The shaded region represents the observations from *Fab12*, circles show the prediction of *EAGLE* and lines show predictions of SA models. Error bars and the thickness of the shaded region represent the  $1\sigma$  uncertainty on the mean and are derived via bootstrap resampling the galaxies in each bin. Unlike *EAGLE*, SA models overpredict the  $H\text{I}$  content of galaxies with  $N \geq 7$  and sSFR below a few  $\times 10^{-11} \text{ yr}^{-1}$ .

at a given sSFR. Indeed, we find that, at the sSFR of  $10^{-12} \text{ yr}^{-1}$ , galaxies with  $N \geq 7$  have on average twice the stellar mass as those with  $N < 3$ , but this alone is not sufficient to explain a difference of a factor  $\sim 3$ –4 in  $f_{H\text{I}}$ . Another possibility is that the  $H\text{I}$ -to- $H_2$  partitioning scheme yields a greater molecular gas fraction in dense environments, owing to the higher pressures in the interstellar and circumgalactic media. To investigate this possibility, we produced plots similar to those of Fig. 2 but for the molecular gas fraction (not shown here for brevity), finding little difference between the two environments. Hence, the only remaining possibility is that the drop in  $f_{H\text{I}}$  in dense environments at a given sSFR is due to an effective drop in the  $H\text{I}$  mass of these galaxies. This indicates that the environment can effectively remove  $H\text{I}$  from galaxy discs while leaving the star-forming gas content largely undisturbed. As already mentioned, such a process is not modelled by SA models, which therefore show a different behaviour at odds with the observations. We stress that these results do not change if we use the *EAGLE* subsample that is unbiased for the environment.

### 3.2 Comparison with GASS observations

*Cat13* used the GASS ‘representative’ sample, which is a complete sample of  $\sim 800$  galaxies with stellar masses  $10 < \log_{10}(M_*/M_\odot) < 11.5$ , to study the environmental effects on the  $H\text{I}$  content of these systems as a function of their stellar mass. The environment in this case is characterized in terms of the virial mass of the host halo to which a galaxy belongs, which has been estimated by Yang et al. (2012) for the SDSS DR7 catalogue using an FoF group finder algorithm. In cosmological simulations, the virial



**Figure 3.**  $\log_{10}(f_{\text{H I}})$  as a function of  $\log_{10}(M_*)$  for all galaxies with  $10 < \log_{10}(M_*/M_\odot) < 11.5$  in EAGLE and in the GASS sample of Cat13. Grey circles represent individual systems in EAGLE. Solid blue and green dashed lines represent, respectively, the average and the median trends in EAGLE, while the blue and the green shaded regions show the average and the median trend for the GASS sample. Error bars are represented by the thickness of the shaded regions and by the size of the circles, and are derived via bootstrapping. In all cases the logarithm is evaluated after the averaging. Dotted lines represent GASS sensitivity and set a lower limit to the H I fractions.

mass of collapsed haloes is often characterized by  $M_{200}$ , the mass contained within a sphere of a radius  $r_{200}$  about a galaxy's centre of potential within which the average matter density is 200 times the critical density of the Universe.

Comparing the prediction of the simulations with the results of Cat13 is complementary to the analysis presented in Section 3.1. On the one hand, Cat13 do not rely on H I stacking to infer the H I fractions in their sample, and they use a clear selection criterion that can be easily applied to our sample of simulated galaxies. On the other hand, the calculation of the host halo virial masses in Yang et al. (2012) is intrinsically different from that implemented in EAGLE: systematics can arise from the different linking lengths adopted to determine the FoF regions, which are defined in redshift-angular space in the SDSS catalogue and in physical space in the simulation, and by the different cosmological parameters adopted (WMAP7 versus Planck). However, we verified visually that the distribution of central and satellite galaxies in the stellar mass–host halo mass plane in EAGLE and in GASS overlap with each other, which is an important starting point for a detailed comparison between the two samples.

We first verify that the EAGLE galaxies have H I properties that are compatible with those of the GASS sample. Fig. 3 shows the H I fractions of each individual EAGLE galaxy with  $10 < \log_{10}(M_*/M_\odot) < 11.5$ , along with the mean and median values computed in different bins of  $M_*$ , and compares them with the mean and median  $f_{\text{H I}}$  derived from the GASS sample. To emulate the GASS selection criteria and enable comparison to Cat13, we applied a lower limit to the  $f_{\text{H I}}$  of our simulated galaxies (shown in Fig. 3 by a dotted line): the H I fraction of systems with  $\log_{10}(M_*/M_\odot) > 10.5$  and  $f_{\text{H I}} < 0.015$  is set to 0.015, and the H I mass of systems with  $\log_{10}(M_*/M_\odot) \leq 10.5$  and  $M_{\text{H I}} < 10^{8.7} M_\odot$  is set to  $10^{8.7} M_\odot$ . With these conditions, the median  $f_{\text{H I}}$  in the EAGLE sample scales with the stellar mass in a way that is consistent with the GASS sample. However, the average  $f_{\text{H I}}$  of simulated galaxies is slightly below the observed values for  $\log_{10}(M_*/M_\odot) > 10.5$ , implying that massive, H I-rich systems in EAGLE are too rare. This was al-

ready noticed by Bahé et al. (2016) and by Crain et al. (2016), who focused on EAGLE centrals and explained the scarcity of H I-rich systems as being due to the presence of spuriously large H I holes in the discs of simulated galaxies. Given that the median H I properties of the simulated and the observed samples are in good agreement with each other for  $M_* > 10^{10} M_\odot$ , we proceed with comparing the environmental trends for the two samples.

Cat13 found that, at a given stellar mass, it is significantly rarer to detect H I in those galaxies that reside in groups, defined as systems with host halo mass greater than  $10^{13} M_\odot$ , with respect to those that reside in lower mass haloes. This result, along with the predictions of EAGLE and of the SA models, is presented in Fig. 4, which shows the fraction of galaxies above the H I detection limit of GASS that reside in groups with host halo mass  $M_{200} < 10^{13} M_\odot$  (left-hand panel) or  $M_{200} > 10^{13} M_\odot$  (second panel from the left). While all SA models predict that the fraction of systems detected in H I should be much larger than the observed values (with the exception of Guo11's model for  $M_{200} < 10^{13} M_\odot$ ), EAGLE is in excellent agreement with the observed trends. This is remarkable, considering that EAGLE has not been calibrated against the H I properties of galaxies. The discrepancy between EAGLE and the SA models is particularly severe for less massive galaxies in groups with  $M_{200} > 10^{13} M_\odot$ , whose H I content is very sensitive to the environment density (as we will see in more detail in Section 4). This discrepancy corroborates the idea that environmental processes are efficient at removing the H I discs of galaxies, and do not affect solely their hot gas reservoirs. As we will see in Section 5.2, environmental mechanisms in EAGLE can fully remove the H I component of a galaxy on short time-scales while leaving its stellar component largely unaffected. We reiterate that these processes are not modelled self-consistently in the SA models, which can explain their different behaviour.

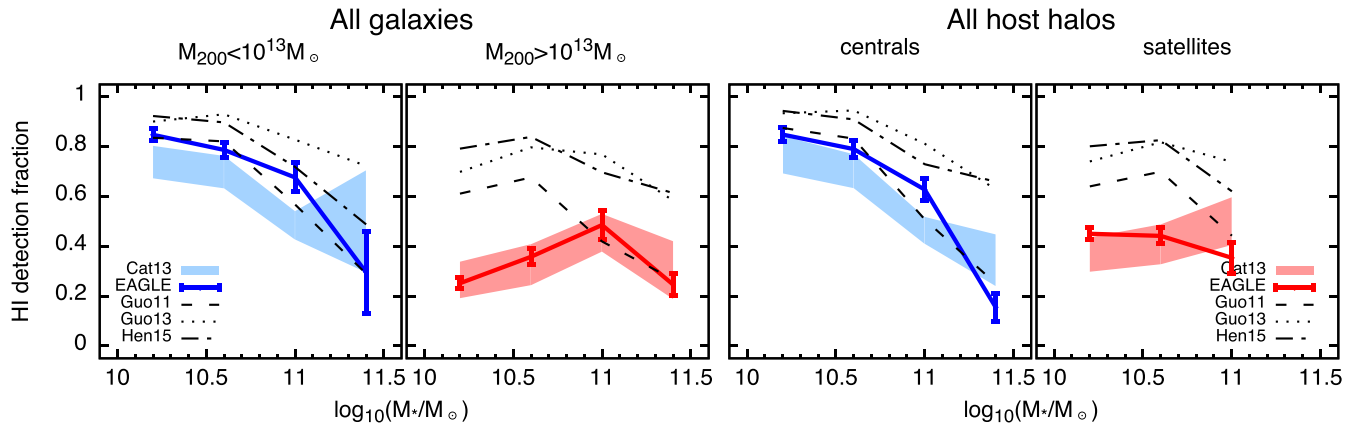
Cat13 found that analogous differences in the H I detection fraction can be seen by splitting the galaxies of the GASS sample into central and satellites, with the former being more frequently detected in H I with respect to the latter at a fixed stellar mass. This result is shown in the two rightmost panels of Fig. 4 and is not surprising, given that satellites live preferentially in massive haloes. Also in this case the SA models predict fractions that are far too large (with the exception of Guo11's model for centrals), while EAGLE is in good agreement with the measurements. Note that, in the last panel of Fig. 4, the last mass bin is missing since there are no satellites in the GASS sample within that mass range (but a few systems are present in the simulations).

There are minor differences between the host halo mass distributions in EAGLE and in the sample of Cat13. As done before, we created an unbiased sample by selectively removing simulated galaxies until the two mass distributions become comparable. The results presented in this section do not change if we use such an unbiased sample rather than the complete one.

#### 4 ANALYSIS OF THE ENVIRONMENTAL TRENDS

In the previous section we have shown that EAGLE reproduces the observed relations between the H I content of galaxies and their local density. The comparison between the simulations and the observations was however limited to the most massive systems ( $M_* > 10^{10} M_\odot$ ). Also, the range of environments probed by the observations is relatively small.

In this section, we extend the relations above down to stellar masses of  $10^9 M_\odot$  – the mass above which we consider galaxies to be well resolved by EAGLE – for a variety of environments. We

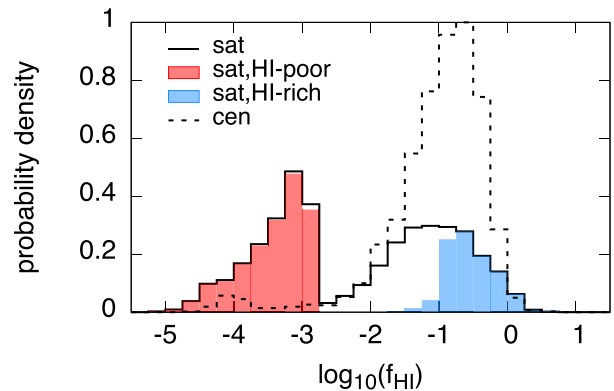


**Figure 4.** Fraction of galaxies above the H I detection limit of GASS that reside in sparse environments ( $M_{200} < 10^{13} M_{\odot}$ , left-hand panel) or dense environments ( $M_{200} > 10^{13} M_{\odot}$ , second panel from the left), or that are classified as centrals (third panel) or satellites (fourth panel) as a function of their stellar mass. The shaded regions represent observations by Cat13, the solid lines show the prediction of EAGLE and the dashed lines show the predictions of the SA models. Error bars in EAGLE and in Cat13 are Poissonian. While SA models overpredict the fraction of systems detectable in H I, EAGLE is in remarkable agreement with the observations, especially for massive host haloes and satellites.

adopt  $M_{200}$  as the estimator of the environment. With respect to the other ‘observational’ environment proxies based on galaxy number densities, such as the quantity  $N$  used in Section 3.1,  $M_{200}$  is a more physically meaningful quantity and has a unique definition, whereas the other estimators are sensitive to the size of the region chosen to compute the density and to the stellar mass/luminosity threshold used. We note that all the observational environment proxies correlate well with  $M_{200}$ , as shown by Haas, Schaye & Jeon-Daniel (2012) and in Appendix A. The use of  $M_{200}$  as a proxy for the environment does not account for all the inhomogeneities and anisotropies in a halo. For instance, Bahé et al. (2013) analysed the GIMIC suite of simulations (Crain et al. 2009) and showed that galaxies which are accreted on to a given halo along filaments experience stronger ram pressure stripping than those accreted from voids. The study of these effects is however beyond the scope of this paper.

We partition the simulated galaxies in five bins of stellar mass, ranging from  $10^9 M_{\odot}$  to  $10^{11.5} M_{\odot}$ , and six bins of  $M_{200}$ , ranging from  $10^{12} M_{\odot}$  to  $10^{14.5} M_{\odot}$ , for a total of 30 bins. We verified that the range of  $M_{200}$  is adequate to sample the majority of the environments to which satellite galaxies in the chosen range of stellar masses belong. We omit from our analysis all central galaxies, i.e. those systems that reside at the centre of each FoF group. This leaves us with a sample of 5549 galaxies in the chosen ranges of stellar mass and  $M_{200}$ . Including central galaxies (an additional 2218 objects) would not alter our results significantly, because satellite galaxies dominate by number in high-mass haloes.

We impose a minimum H I mass of  $M_{\text{H I, min}} = 0.752 \times 1.82 \times 10^6 M_{\odot}$  to all our systems. This value corresponds to the mass of a single gas particle multiplied by the primordial hydrogen abundance and can be regarded as the resolution limit for the (total) hydrogen mass of our simulated galaxies. Also, we define H I-rich and H I-poor galaxies at given  $M_*$  in our sample according to whether they occupy the highest or the lowest quartile in the H I mass distribution. These H I quartiles are first computed in bins of  $\Delta \log_{10} M_* = 0.25$ , and then fit with a second-order polynomial to yield  $M_{\text{H I}}(M_*)$  quartile relations. In practice, this definition implies that only those systems with  $M_{\text{H I}} = M_{\text{H I, min}}$  are H I poor, while most H I-rich galaxies have  $M_{\text{H I}} > 0.1 M_*$ . In Fig. 5 we plot the distribution of  $\log_{10}(f_{\text{H I}})$  for our sample of satellites (solid histogram) and compare it to that derived for centrals (dashed histogram). The two distributions differ significantly. Satellite systems show a bimodal distribution with a



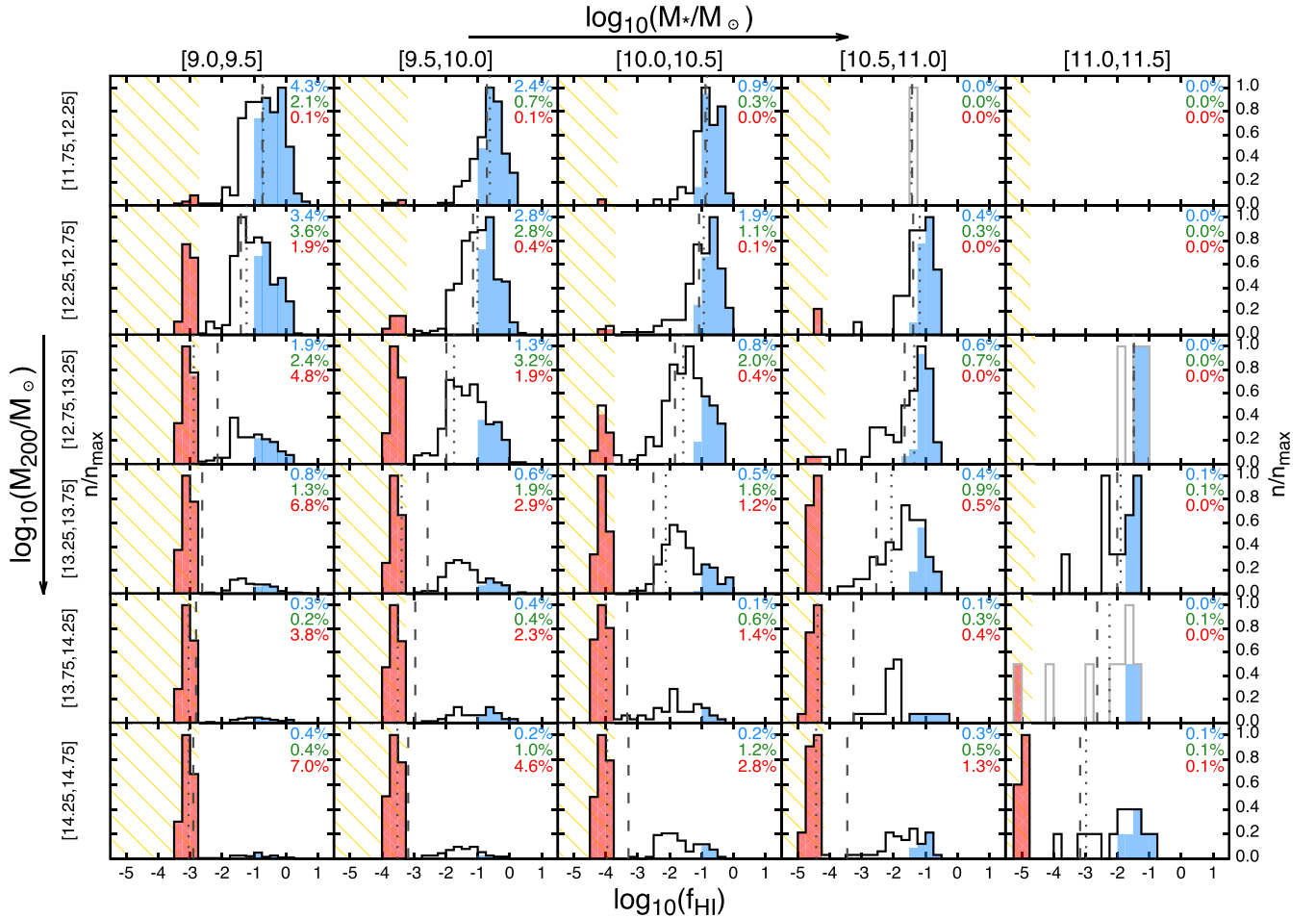
**Figure 5.** Distribution of  $\log_{10}(f_{\text{H I}})$  for satellites (solid histogram) and centrals (dashed histogram) in EAGLE. All galaxies have stellar masses between  $10^9$  and  $10^{11.5} M_{\odot}$  and reside in haloes with  $M_{200}$  between  $10^{11.75}$  and  $10^{14.75} M_{\odot}$ . The blue-shaded and red-shaded regions represent H I-rich (top quartile in  $M_{\text{H I}}(M_*)$ ) and H I-poor (bottom quartile) satellites, respectively. The sharp upturn at  $\log_{10}(M_{\text{H I}}/M_*) \simeq -3$  is due to the minimum H I mass assigned to galaxies (corresponding to a single particle). The central and satellite distributions differ markedly, with the latter showing a significant fraction of H I-poor systems.

prominent peak around  $\log_{10}(M_{\text{H I}}/M_*) = -3$  that is absent from the distribution of centrals. This peak is almost completely occupied by H I-poor systems (red-shaded histogram). Note that this peak is artificial, being due to the minimum H I mass assigned to galaxies: H I fractions cannot be lower than  $M_{\text{H I, min}}/M_*$ , and this produces the sharp upper edge around  $\log_{10}(M_{\text{H I}}/M_*) = -3$ . Nonetheless, the difference between the two distributions highlights the dramatic impact of the environment on the H I content of the satellite galaxy population. Centrals and satellites in our sample have a similar stellar mass distribution (not shown here). We now explore how satellites with different stellar mass living in host haloes with different masses contribute to the global  $f_{\text{H I}}$  distribution.

#### 4.1 H I fractions as a function of $M_*$ and $M_{200}$

As we show below, a simple study of how the typical H I content of galaxies depends on  $M_*$  and on  $M_{200}$  would not give a complete





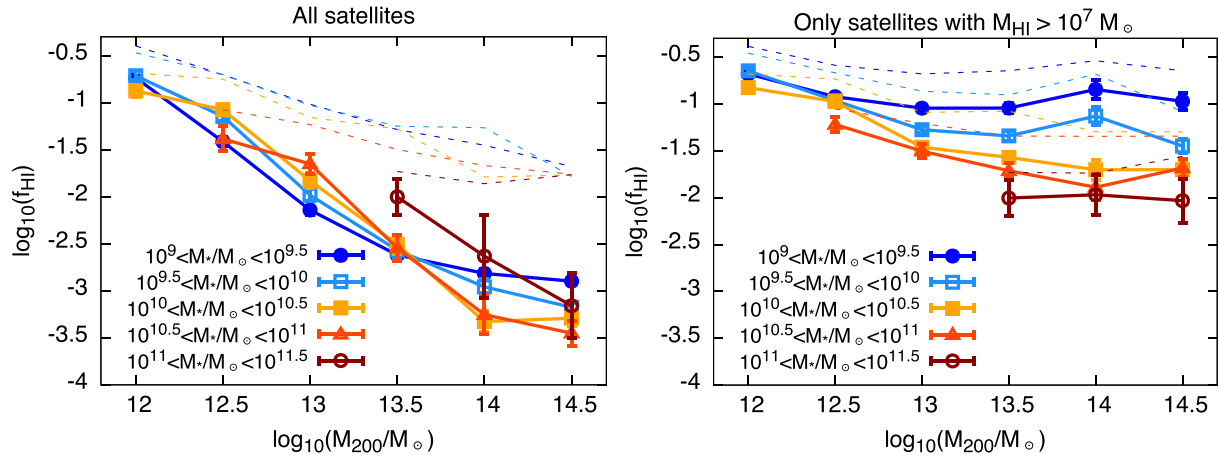
**Figure 6.** Distributions of  $\log_{10}(f_{\text{HI}})$  in different bins of stellar mass and  $M_{200}$  for satellites in EAGLE, after imposing a minimum H I mass of  $1.36 \times 10^6 M_{\odot}$  (corresponding to a single particle). Each distribution is normalized to its peak value. Stellar masses increase from left to right (as indicated on top of the figure),  $M_{200}$  increases from top to bottom (as indicated to the left). Grey histograms are used for bins with fewer than 10 galaxies. The dashed and dotted vertical lines represent, respectively, the mean and the median of each distribution. The striped region in the left-hand part of each panel is resolution limited. The top-right of each panel lists the percentage of galaxies that are H I-rich (top quartile in  $M_{\text{HI}}(M_*)$ , blue), H I-poor (bottom quartile, red) or in between the two (green), normalized to the whole sample. The fraction of H I-poor satellites increases with increasing  $M_{200}$  at fixed  $M_*$ , and decreases with increasing  $M_*$  at fixed  $M_{200}$ .

picture of how these quantities correlate with the gas properties. In Fig. 6 we show the distribution of  $\log_{10}(f_{\text{HI}})$  for the EAGLE satellites in bins of  $M_*$  and  $M_{200}$ . Stellar mass increases from left to right,  $M_{200}$  increases from top to bottom (as indicated to the left). Note that in the most diffuse environment considered in our sample (top row)  $f_{\text{HI}}$  follows a roughly lognormal distribution. This has been reported by Cortese et al. (2011) for observational measurements and implies that, if we aim to characterize  $f_{\text{HI}}$  in terms of the moments of its distribution, it would be more appropriate to focus on the distribution of its logarithm, as we do in this analysis.

The effects of the environment on the H I content of satellites can be studied by moving vertically along the panels of Fig. 6, i.e. by analysing how the distribution changes as a function of  $M_{200}$  at a given  $M_*$ . We can identify a main and a secondary environmental effect. The main effect is that, when  $M_{200}$  becomes sufficiently large, the distribution of  $\log_{10}(f_{\text{HI}})$  begins to show a bi-modality as H I-poor systems – represented as red-shaded histograms in Fig. 6 – emerge as a secondary peak besides the lognormal distribution visible for the most diffuse environment. This peak becomes more prominent as  $M_{200}$  increases, and for the densest regions probed by EAGLE dominates the galaxy number density (see fractions on the

top right-hand corner of each panel). A secondary effect is that, as  $M_{200}$  increases, the portion of the distribution that is not H I poor moves slowly towards smaller values of  $f_{\text{HI}}$ . This is more evident for  $9.5 < \log_{10}(M_*/M_{\odot}) < 10.5$ .

The fraction of H I-poor galaxies increases with increasing  $M_{200}$  at given  $M_*$ , and decreases with increasing  $M_*$  at given  $M_{200}$ . Thus the precise  $M_{200}$  at which H I-poor systems appear depends on the stellar mass. For instance, H I-poor galaxies with  $9 < \log_{10}(M_*/M_{\odot}) < 9.5$  are already present around  $M_{200} = 10^{12.5} M_{\odot}$ , while those with  $10.5 < \log_{10}(M_*/M_{\odot}) < 11$  emerge only above  $M_{200} \simeq 10^{13.5} M_{\odot}$ . This trend can be interpreted as follows: at a given  $M_{200}$ , ram pressure stripping, tidal interactions and other environmental processes influence lower mass galaxies more efficiently, because the gravitational restoring force is smaller and their H I is less tightly bound to the system. Conversely, the number of H I-rich galaxies, represented by the blue-shaded part of the histograms, decreases while moving from low- to high-mass host haloes. On the one hand, this is caused by the corresponding growth of H I-poor systems with  $M_{200}$ . On the other hand, the H I-rich part of the distribution moves on average towards lower  $f_{\text{HI}}$  as  $M_{200}$  increases. These two effects combine to produce a significant



**Figure 7.** Average  $f_{\text{HI}}$  predicted by EAGLE as a function of the host halo mass  $M_{200}$ , imposing a minimum H I mass of  $1.36 \times 10^6 M_{\odot}$  (corresponding to a single particle). Different lines represent satellites with different stellar masses. The left-hand panel shows the results for all satellites, the right-hand panel shows only satellites with  $M_{\text{HI}} > 10^7 M_{\odot}$ . Solid lines show  $\langle \log_{10}(f_{\text{HI}}) \rangle$ , dashed lines show  $\log_{10}(\langle f_{\text{HI}} \rangle)$ . Error bars represent the  $1\sigma$  variance on the average and are derived via bootstrap resampling the galaxies in each bin. The trends of the average  $f_{\text{HI}}$  with the host halo mass depend on how the averaging is performed and on whether or not H I-poor systems are included.

drop in the average H I mass fractions for galaxies with increasing  $M_{200}$ . This is clearly illustrated by the shift of the vertical dashed lines, which represent the mean (dashed) and the median (dotted)  $\log_{10}(f_{\text{HI}})$ , from the top to the bottom panels of Fig. 6. We also notice that, in line with the observational findings of Cat13, at any given  $M_{200}$ , the H I-rich part of the distribution is truncated at lower  $f_{\text{HI}}$  when moving towards more massive satellites.

We reiterate that systems corresponding to the H I-poor peak have their H I content fixed at  $1.36 \times 10^6 M_{\odot}$ , the (hydrogen) mass resolution of the simulation. As the true H I mass of these galaxies might be anything between zero and this value, the red peaks in the histograms are artificial and the true shape of the distribution below the resolution limit, represented by the yellow-dashed region on left-hand side of each panel of Fig. 6, cannot be predicted by the simulation. However, regardless of their true H I mass, these galaxies would still belong to the first quartile of the H I mass distribution at any given  $M_*$  and therefore would be classified as H I poor. We stress that these results do not change significantly when central galaxies are included in our sample. Centrals preferentially occupy the most diffuse environments and do not show a prominent peak of H I-poor systems (see Fig. 5), just like satellite galaxies in diffuse environments, and they therefore do not alter the overall picture described here.

We now show that the ‘average’  $f_{\text{HI}}$  as a function of environment and stellar mass depends both on how the averaging is performed, and on which galaxies are selected. The left-hand panel of Fig. 7 shows the trend of the average  $f_{\text{HI}}$  with  $M_{200}$  for EAGLE satellites in five bins of stellar mass. Here the averaging is computed in two ways: the solid lines show the mean of  $\log_{10}(f_{\text{HI}})$ , while the dashed lines show the logarithm of the mean  $f_{\text{HI}}$ . In the first case, it appears that the host halo mass determines the average H I content of a galaxy:  $f_{\text{HI}}$  decreases by 2–2.5 orders of magnitude when increasing from  $M_{200} \sim 10^{12}$  to  $\sim 10^{14.5} M_{\odot}$  and variations in  $f_{\text{HI}}$  at fixed  $M_{200}$  as a function of stellar mass are minimal. Note that the slope of this relation may be underestimated, given that only an upper limit on  $M_{\text{HI}}/M_*$  is available for the H I-poor systems. The cause of this trend can be investigated by following the vertical dashed lines in Fig. 6. At any fixed  $M_*$ , when  $M_{200}$  is increased the fraction of H I-

poor galaxies increases and the distribution shifts slowly towards lower  $f_{\text{HI}}$ , leading to a net reduction in the average value. At a fixed  $M_{200}$ , the fraction of H I-poor galaxies drops when increasing  $M_*$ , but simultaneously the H I-rich tail of the distribution is truncated, producing little net variation in the mean. This is particularly evident for systems at  $M_{200} \sim 10^{12.5} M_{\odot}$ . The dashed lines in the left-hand panel of Fig. 7 follow instead a much shallower slope, which is not surprising given that H I-poor systems heavily affect the logarithmic average. In this case it would appear that both  $M_*$  and  $M_{200}$  play a role in establishing the mean H I fractions.

A different conclusion would be reached if we were unaware of the presence of galaxies that are almost completely devoid of H I, which might be the case for an H I-selected sample of galaxies. This is shown in the right-hand panel of Fig. 7, where we consider a scenario where only galaxies with  $M_{\text{HI}} > 10^7 M_{\odot}$  are considered. Here, all trends with the environment are reduced with respect to the full sample, and stellar mass plays a more dominant role in setting the average H I content of galaxies regardless the method adopted to compute the mean. Obviously, discarding the H I-depleted systems from the analysis leads to a systematic underestimate of the effect of the environment on the global H I content of galaxies. This is because the environment seems to control mainly whether or not a galaxy has any H I rather than inducing a continuous trend, suggesting that the H I removal happens on short time-scales. We will discuss this further in Section 5.2.

In Appendix B we show the good numerical convergence of the  $f_{\text{HI}}-M_{200}$  relation for the EAGLE satellites, which suggests that the resolution of the Ref-L100N1504 run is adequate to model the physics of the environmental processes. For in-depth tests of both numerical resolution and systematic uncertainties related to the partitioning of hydrogen into the ionic/atomic/molecular phases, we redirect the reader to Crain et al. (2016).

## 4.2 H I properties as a function of the groupcentric distance

Galaxies located in the proximity of the cluster centres have, on average, less H I than those located at the periphery of these systems or than field galaxies (Giovanelli & Haynes 1985; Haynes &

Giovanelli 1986; Solanes et al. 2001; Gavazzi et al. 2006). Classically, this is quantified in terms of ‘H I deficiency’ ( $\text{def}_{\text{H I}}$ ; Haynes & Giovanelli 1984), defined for a given galaxy as

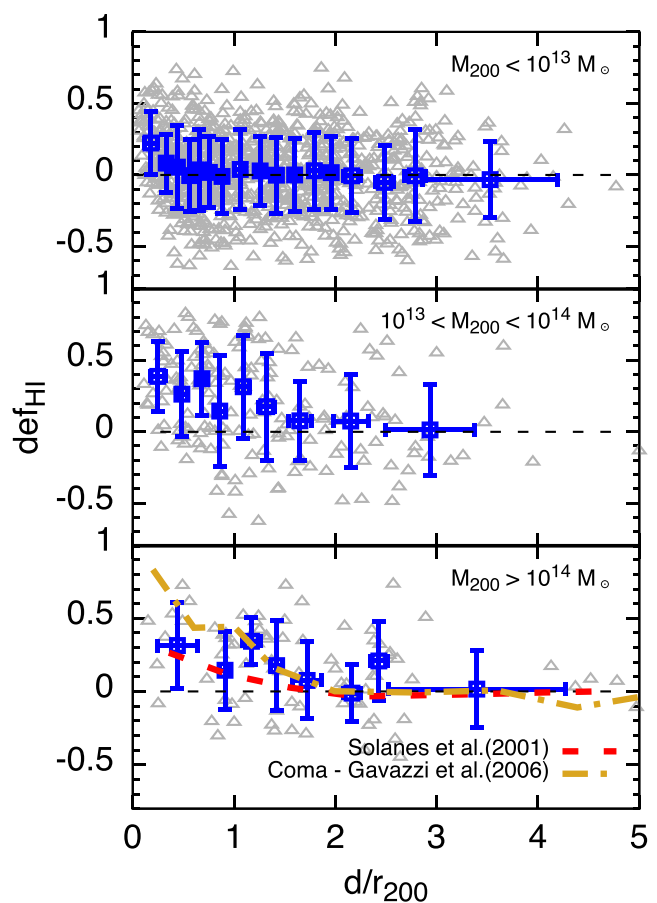
$$\text{def}_{\text{H I}} = \log_{10}(M_{\text{H I,ref}}) - \log_{10}(M_{\text{H I,obs}}), \quad (1)$$

where  $M_{\text{H I,obs}}$  is the total H I mass of the observed system and  $M_{\text{H I,ref}}$  is the typical H I mass for a reference sample of isolated galaxies with stellar properties (i.e. Hubble type and diameter) similar to those of the observed galaxy. In nearby clusters, the mean  $\text{def}_{\text{H I}}$  remains close to 0 for galaxies located beyond 1–2 virial radii from the cluster centre, and it increases to  $\approx 0.5$  in the innermost regions.

We now verify whether or not satellites in EAGLE follow a similar trend. In the Ref-L100N1504 run at  $z = 0$ , the number of FoF groups with virial mass in the cluster range is small: there are seven groups with  $M_{200} > 10^{14} M_{\odot}$ , and none with  $M_{200} > 5 \times 10^{14} M_{\odot}$ . Thus, Ref-L100N1504 does not contain a galaxy cluster as massive as Virgo or Coma. However, it remains interesting to check if the observed trend is reproduced also at lower host halo masses. Another concern is that the original definition of  $\text{def}_{\text{H I}}$  uses the galaxy Hubble type, which is not easily determined in the simulation. Instead, we calibrate  $M_{\text{H I,ref}}$  on the simulated *centrals* via the following procedure. We select galaxies with  $M_{*} > 10^9 M_{\odot}$  and  $M_{\text{H I}} > 6.8 \times 10^8 M_{\odot}$  in Ref-L100N1504. This threshold in H I mass corresponds to having at least 500 gas particles in each system, ensuring that the ISM structure is sampled sufficiently, and is similar to the H I mass sensitivity for the observations in the Coma Cluster by Gavazzi et al. (2006). We focus on the centrals of this sample, for which we derive the median  $M_{\text{H I}}$  as a function of  $M_{*}$  using bins of 0.2 dex in  $M_{*}$ . This relation is then fit with a polynomial, which we use to determine  $M_{\text{H I,ref}}$  given the stellar mass of the satellite.

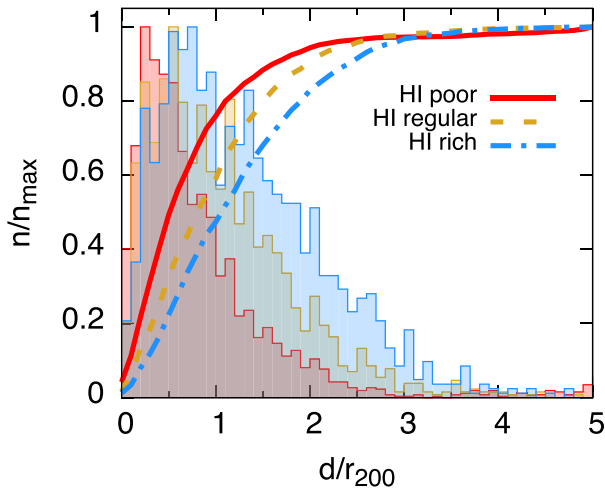
In Fig. 8 we show our calibrated  $\text{def}_{\text{H I}}$  for satellites with  $M_{\text{H I}} > 6.8 \times 10^8 M_{\odot}$  in the Ref-L100N1504 run as a function of their distance from the group centre. The three panels show galaxies located in haloes with  $M_{200} < 10^{13} M_{\odot}$  (top panel),  $M_{200} > 10^{14} M_{\odot}$  (bottom panel) or in between the two (central panel). In the least massive halo bin, the mean  $\text{def}_{\text{H I}}$  tends to zero for  $d > 0.5 r_{200}$ , and it increases only in the innermost region. In more massive haloes, the mean  $\text{def}_{\text{H I}}$  departs from zero already around  $d = 2 r_{200}$ , reaching a value of  $\approx 0.4$  at the group centre. We found that the galaxies that contribute to the increase of  $\text{def}_{\text{H I}}$  are those with  $M_{*} > 10^{10} M_{\odot}$ , while less massive systems do not show significant H I deficiency. At first glance, this may appear at odds with the results of Section 4.1, where we showed that satellites with  $M_{*} < 10^{10} M_{\odot}$  are more sensitive to the environment than more massive systems (see Fig. 6). Clearly, this is due to the fact that H I-poor systems – being undetectable in H I – do not contribute to the H I deficiency. In fact, if we discarded all the H I-poor satellites from Fig. 6, the only remaining trend with  $M_{200}$  would be a global shift of  $f_{\text{H I}}$  to lower values, which is more evident for  $M_{*} > 10^{10} M_{\odot}$ . In the bottom panel of Fig. 8 we also show the mean  $\text{def}_{\text{H I}}$  measured by Gavazzi et al. (2006) for galaxies in the Coma Cluster (for which we assume  $r_{200} = 2.2$  Mpc) and the mean  $\text{def}_{\text{H I}}$  derived by Solanes et al. (2001) in 18 other nearby clusters. Although we did not attempt to mimic precisely all the selection criteria and biases present in the observations, it can be clearly seen that the prediction of EAGLE is in good agreement with the data. In general, EAGLE is compatible with the idea that galaxies located within  $\sim 2 r_{200}$  from the centre of groups with  $M_{200} > 10^{13} M_{\odot}$  are more H I deficient than those in the field.

We reiterate that the analysis of  $\text{def}_{\text{H I}}$  is based on galaxies with detectable H I masses. This gives an incomplete picture of how the environment affects satellites as a function of the groupcentric



**Figure 8.** H I deficiency (equation 1) as a function of the distance to the centre of the group (normalized by  $r_{200}$ ) for satellites with  $M_{*} > 10^9 M_{\odot}$  and  $M_{\text{H I}} > 6.8 \times 10^8 M_{\odot}$  in the Ref-L100N1504 run. The three panels show, from top to bottom, satellites located in haloes with  $M_{200} < 10^{13} M_{\odot}$ ,  $10^{13} < M_{200} < 10^{14} M_{\odot}$  and  $M_{200} > 10^{14} M_{\odot}$ . Grey triangles show individual systems, squares with error bars show averages and standard deviations derived for every 50 (top), 25 (central) and 10 (bottom) objects. In the bottom panel we also report the mean H I deficiencies found by Solanes et al. (2001, dashed line) in a sample of 18 nearby galaxy clusters and by Gavazzi et al. (2006, dot-dashed line) in the Coma Cluster. Both observations and the simulation indicate that the deficiency increases towards the group centres. The trend is stronger at higher  $M_{200}$ .

distance, given that, as we have seen in Section 4.1, satellites in the most massive haloes tend to be H I poor (i.e. their H I mass is lower than the particle mass resolution of the simulation). To offer a more complete perspective, we show in Fig. 9 the distributions of  $d/r_{200}$  for all satellites in our simulated galaxy sample, distinguishing between galaxies that are H I poor (red histogram), H I rich (blue histogram) or in between (‘H I regular’, yellow histogram). Clearly, these three categories of systems have a different spatial distribution within their groups, with the H I-poor and H I-rich satellites being the most and least concentrated, respectively. This is in line with the observational findings of Hess & Wilcots (2013) based on ALFALFA data, and can be quantified by looking at the cumulative distributions (solid lines in Fig. 9): 76 per cent of the H I-poor satellites are located within the virial radius, compared to 47 per cent for the H I-rich systems. This discrepancy is more severe at  $0.5 r_{200}$ , where these fractions become 50 and 22 per cent, respectively. Here we do not show the  $d/r_{200}$  distribution for different ranges of stellar



**Figure 9.** Distances of satellite galaxies (with  $M_* > 10^9 M_\odot$ ) from their group centre, normalized to  $r_{200}$ , in the Ref-L100N1504 run. The solid, dashed and dot-dashed lines show H I-poor (bottom quartile in  $M_{H I}(M_*)$ ), H I-regular (second and third quartiles) and H I-rich (top quartile) satellites, respectively. The thin lines show the number distributions normalized to their peak values, the thick lines show the cumulative distributions. H I-poor satellites reside preferentially in the innermost regions of their group, whereas H I-rich systems are more sparsely distributed.

mass and host halo mass, as we did not find a significant trend with these two quantities.

A similar – but more pronounced – segregation between H I-rich and H I-poor satellites is observed in the Local Group, where the dwarf galaxies located within a virial radius from either the Milky Way or M31 have virtually no cold gas (with the notable exception of the Magellanic Clouds), while those farther away are all bright in H I (Grcevich & Putman 2009; Spekkens et al. 2014). Unfortunately the mass resolution of EAGLE does not enable us to probe the dwarf regime. Zoom-in cosmological hydrodynamical simulations, such as the APOSTLE runs (Sawala et al. 2016; Fattahi et al. 2016), constitute an excellent tool to extend the analysis presented here to galaxies with  $M_* < 10^9 M_\odot$ .

## 5 THE PHYSICAL ORIGIN OF THE ENVIRONMENTAL TRENDS

In this section we discuss which environmental effects cause a satellite galaxy in the simulation to lose gas and become H I poor. We focus on those processes that act on a galaxy’s ISM and can thus *directly* remove H I from a system. We note that other environmental processes can indirectly alter the H I content of a galaxy: an example is the ‘starvation’ mechanism (e.g. Bekki, Couch & Shioya 2002), where the ram pressure or tidal stripping of a galaxy’s hot gas reservoir inhibits further gas accretion on to the disc, with long-term consequences to the cold gas content of the system. We distinguish between three different mechanisms: ram pressure stripping by the galaxy’s motion relative to the intragroup medium (IGM), tidal stripping by the host halo and high-speed satellite–satellite interactions. We first verify whether or not these mechanisms are at work in the simulation at  $z = 0$ , which galaxies are currently affected by them, where these systems are located and what is their H I morphology. Then, we focus on the H I-poor satellites at  $z = 0$

and track their H I content back in time to relate their gas loss to one or more of these mechanisms.

### 5.1 Environmental effects at $z = 0$

We focus on satellite galaxies with  $M_* > 10^9 M_\odot$  and  $M_{H I} > 6.8 \times 10^8 M_\odot$  in Ref-L100N1504. As in Section 4.2, this threshold in H I mass ensures that the satellite’s ISM is adequately sampled. We find 1404 systems that meet these criteria. For each system, we estimate the environmental effects that it is currently experiencing as follows.

(i) *Ram pressure stripping.* We use the classical formula of Gunn & Gott (1972) to establish whether the pressure exerted by the IGM on to the ISM of a galaxy suffices to overcome its gravitational restoring force at a given radius  $R$ . This happens when

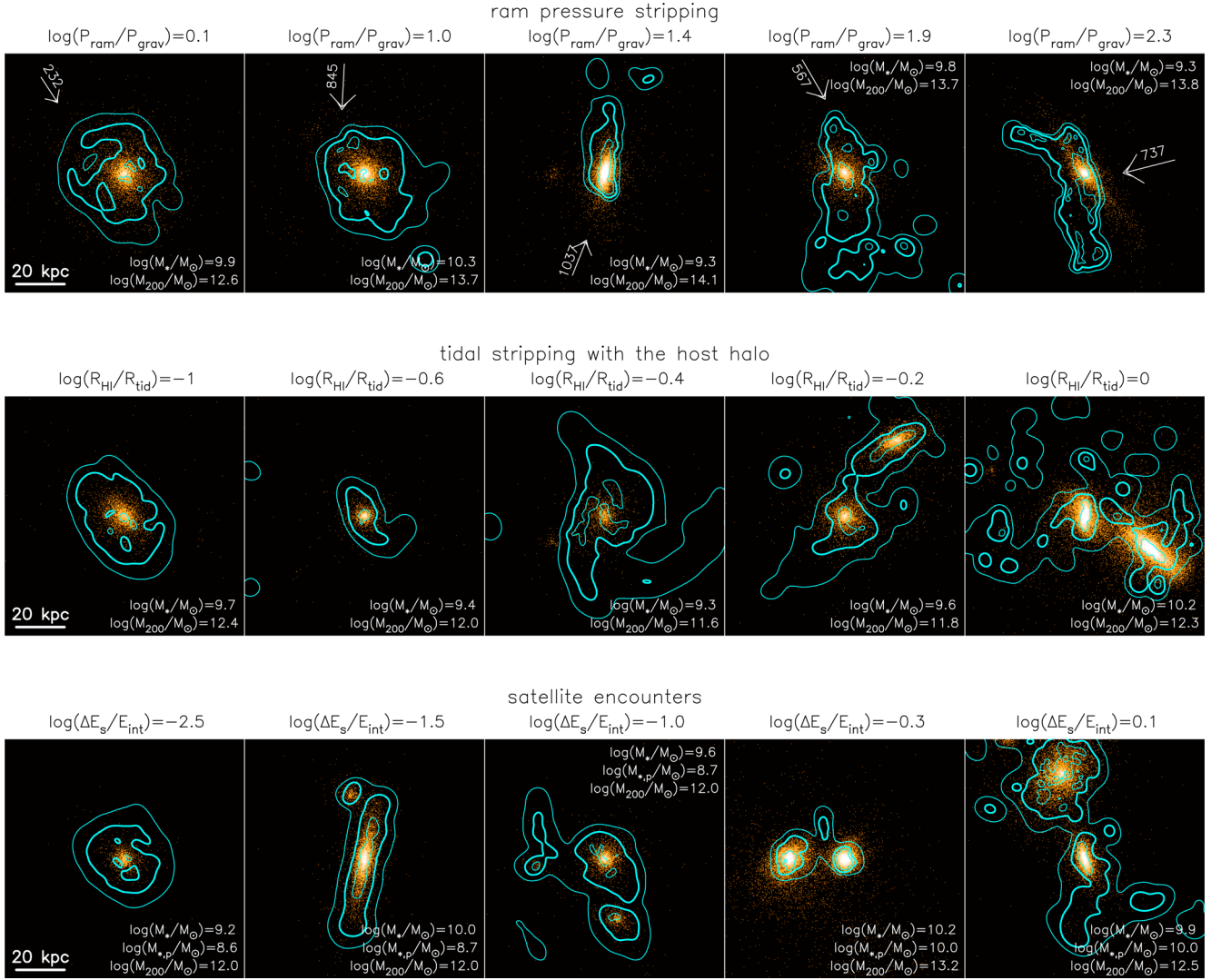
$$\rho v^2 > \left. \frac{\partial \Phi(R, z)}{\partial z} \right|_{z=0} \Sigma_{\text{ISM}}(R), \quad (2)$$

where  $\rho$  is the IGM density,  $v$  is the relative velocity between the satellite and the surrounding IGM,  $\Phi$  is the gravitational potential,  $\partial \Phi(R, z)/\partial z|_{z=0}$  is the gravitational acceleration towards the mid-plane at radius  $R$  and  $\Sigma_{\text{ISM}}(R)$  is the ISM surface density at this radius. Equation (2) assumes that the IGM flows perpendicular to the galaxy disc, but we apply it regardless of the ‘wind’ direction in order to derive a rough estimate of the ram pressure. Typically, the wind direction has little influence on the gas loss (Roediger & Brüggen 2007). The mean density and velocity of the IGM are evaluated as mass-weighted quantities over the nearest 500 particles about the satellite’s centre which are not members of any gravitationally bound subhalo (except for the host subhalo). This allows us to focus on the intergalactic material that surrounds the system – at typical distances ranging from 15 to 100 kpc – and avoid contamination by other satellites, whose environmental influence is studied separately (see below). The number of gas particles adopted ensures that the IGM around the satellites is well sampled. We tried to vary this number by a factor of a few and found no significant difference in our results.

We compute the right-hand-side of equation (2) at the galactocentric radius  $R_{H I}$ , the radius beyond which the H I column density drops below  $1 M_\odot \text{pc}^{-2}$ . At this radius, we expect most hydrogen to be in atomic form, thus  $\Sigma_{\text{ISM}}(R_{H I}) \simeq \Sigma_{H I}(R_{H I})/X_H = 1/X_H M_\odot \text{pc}^{-2}$ , with  $X_H \simeq 0.752$ . Bahé et al. (2016) have shown that centrals in EAGLE follow closely the observed  $M_{H I}$ – $R_{H I}$  relation of Broeils & Rhee (1997), thus we infer the value of  $R_{H I}$  from the total H I mass of our systems. In practice,  $R_{H I}$  is the radius where the H I surface density would be  $1 M_\odot \text{pc}^{-2}$  if the system were unperturbed. In order to evaluate the restoring acceleration, we first rotate our systems to a face-on view by projecting all particles to the reference frame given by the eigenvectors of the system’s inertia tensor. The latter is derived for star particles within  $R_{H I}$ . We checked visually that, with this method, galaxies are aligned face-on. Then, we approximate the partial derivative in equation (2) as  $[\Phi(R_{H I}, 2\epsilon) - \Phi(R_{H I}, 0)]/2\epsilon$ , where  $\epsilon = 2.66 \text{ kpc}$  is the Plummer-equivalent gravitational softening length. We verified that a different choice of  $\epsilon$  has little impact on our results. In our case, the condition expressed by equation (2) implies that ram pressure by the IGM is capable of stripping the galaxy’s H I at the location where the ISM has a surface density of  $1/X_H M_\odot \text{pc}^{-2}$ . We calculate the ratio between the left-hand side ( $P_{\text{ram}}$ ) and the right-hand side ( $P_{\text{grav}}$ ) of equation (2) for all satellites in the simulation at  $z = 0$ .

The top row of Fig. 10 shows the stellar and H I maps for five representative satellites where  $\log_{10}(P_{\text{ram}}/P_{\text{grav}})$  is respectively 0.1,





**Figure 10.** Total stellar and H I maps for a sample of EAGLE satellites extracted from the Ref-L100N1504 run at  $z = 0$ . All maps are on the same scale. Background colours represent stellar surface density, contours represent H I column densities of  $10^{19}$ ,  $10^{20}$  (thickest contour) and  $10^{21} \text{ cm}^{-2}$ . From the left to the right, we show systems that are increasingly affected by ram pressure by the IGM (top row), tidal interactions with the host halo (central row) and interactions with other satellites (bottom row). The stellar mass of the satellite  $M_*$  and the group virial mass  $M_{200}$  are reported on each panel. The arrows in the top panels show the direction of the IGM motion relative to the satellite, and its velocity in  $\text{km s}^{-1}$ . The stellar mass of the perturber satellite,  $M_{*,p}$ , is reported in the bottom panel.

1, 1.4, 1.9 and 2.3. In all cases, we verified that ram pressure effects are dominant with respect to the other environmental mechanisms studied here (i.e. the latter are well below their assumed threshold values, see below). Clearly, the H I morphology becomes more and more disturbed when moving towards the rightmost panels, indicating that equation (2) can be used to pinpoint effectively those systems for which ram pressure is significant. While the first system has an undisturbed H I morphology, the second begins to show H I contours that are more elongated in the direction of the wind flow. The next two systems show clear head–tail H I morphologies, but the most significant case is the last one, where the H I bends spectacularly in the direction of the IGM flow. A similar feature is observed in NGC 4405 in the Virgo Cluster (Chung et al. 2009). In this simulated galaxy, the ram pressure is so strong that even the innermost H I contour is significantly displaced from the stellar disc. Note that generally the direction of the wind is consistent with the galaxy’s H I morphology.

(ii) *Tidal stripping.* The tidal radius,  $r_t$ , for a satellite that moves in a circular orbit around a spherical host halo can be approximated as

$$r_t \approx \left[ \frac{m}{3M(r)} \right]^{\frac{1}{3}} r, \quad (3)$$

where  $m$  is the total mass of the satellite and  $M(r)$  is the total (baryonic+dark) matter mass within the orbit radius  $r$  (e.g. Binney & Tremaine 2008, p. 681). We expect that material beyond the tidal radius can be stripped from the satellites because of the gravitational pull of the main halo. In the more realistic case of eccentric orbits equation (3) is valid only when the satellite is close to the orbit’s pericentre: in this case  $r$  is the pericentric radius. The tidal radius is not a well-defined quantity at any point of a general orbit. As we are interested in deriving a proxy for the ongoing tidal effects, we make the crude assumption that all satellites are on circular orbits and evaluate  $r_t$  by using equation (3), with  $r$  being the current distance

between the satellite and the potential minimum of its host halo. For consistency with the analysis of the ram pressure effects, we compare the tidal radius with  $R_{\text{H1}}$  (see above).

The central row of Fig. 10 shows the stellar and H I maps for five representative satellites where  $\log_{10}(R_{\text{H1}}/r_t)$  is, respectively,  $-1$ ,  $-0.6$ ,  $-0.4$ ,  $-0.2$  and  $0$ . In all five cases we have verified that the other environmental effects are not strong, so that we can focus on tidal effects alone. While the first system has an unperturbed H I morphology, all the others show some degree of disturbance in their stellar and, especially, in their H I components. In the last two cases, the satellite is close to the central galaxy of the group and this distorts their H I reservoirs into spectacular filaments that, at the column density of  $10^{19} \text{ cm}^{-2}$ , encompass both galaxies. In general, it is not straightforward to predict the fate of the H I that is tidally perturbed. However, we feel confident in using equation (3) to identify systems that are tidally perturbed in the simulation.

(iii) *Satellite encounters.* High-speed encounters between satellite galaxies can produce tidal shocks that dynamically heat the systems. The amount of heat  $E_s$  that an extended satellite of total mass  $M_s$  gains during an encounter with a point-like system of total mass  $M_p$  can be computed via the impulsive approximation as

$$E_s \approx \frac{4}{3} G^2 M_s \left( \frac{M_p}{v} \right)^2 \frac{\langle r^2 \rangle}{b^4}, \quad (4)$$

where  $v$  is the relative velocity between the two objects,  $b$  is the impact parameter and  $\langle r^2 \rangle$  is the mass-weighted mean square radius ( $\sum_i m_i r_i^2 / \sum_i m_i$ ) of the extended system (e.g. Binney & Tremaine 2008, p. 660). Equation (4) can be used to predict the outcome of a high-speed encounter between a satellite pair, but it is not generally applicable to the ongoing mutual interaction between any given satellite pair. Also, equation (4) is valid in the limit  $b \gg r_s$ , with  $r_s$  being the typical size of the extended system. As for the other mechanisms studied, we simply use equation (4) as a proxy for the ongoing effect of the interaction between satellites, and then we show that those systems with  $E_s$  above a given threshold actually have a disturbed H I morphology. We evaluate  $E_s$  by using equation (4) where we substitute the current distance between the satellite pair for the impact parameter  $b$ . In practice, for each satellite in our sample we evaluate the maximum  $E_s$  by considering all possible ‘perturber’ subhaloes belonging to the satellite’s group (excluding the central subhalo), and we compare this value with the total (kinetic+potential) internal energy  $E_{\text{int}}$  of the satellite. We stress that the perturbers considered are all the satellite subhaloes identified by the SUBFIND algorithm, and are not limited to those with  $M_* > 10^9 M_\odot$ . We verified that the results presented below do not change significantly if, instead of the maximum  $E_s$ , we use the sum of  $E_s$  over all the perturbers. This suggests that the role of multiple simultaneous interactions is minor, i.e. a single perturber typically dominates the dynamical heating of a system.

The bottom row of Fig. 10 shows five examples of systems that are increasingly affected by this environmental mechanism, while ram pressure and tidal interactions with the host halo are negligible. The first system is unperturbed ( $\log_{10}(E_s/E_{\text{int}}) = -2.5$ ), but the others have H I morphologies with an increasing complexity. The third system is particularly interesting, as it consists of three interacting satellites surrounded by a common H I envelope at a column density of  $10^{19} \text{ cm}^{-2}$ . Note that while the interaction between these galaxies is evident from their H I morphology, their stellar components are only marginally disturbed, similar to what is observed in the M81 group (Yun et al. 1994). This is also the case for the last two systems,

reconfirming the importance of H I as a tracer of the environmental mechanisms.

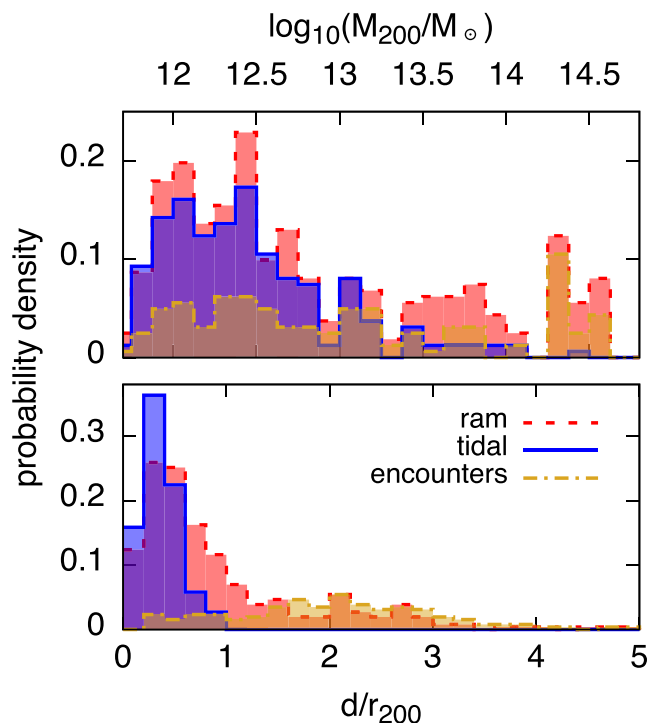
Our findings indicate that all of the three mechanisms considered can influence the H I discs of satellites with  $M_* > 10^9 M_\odot$  and  $M_{\text{H1}} > 6.8 \times 10^8 M_\odot$  at  $z = 0$ . The maps shown in Fig. 10 help us in establishing ‘ad hoc’ thresholds above which the influence of each mechanism is significant: ram pressure begins to affect significantly the H I morphology of satellites when  $\log_{10}(P_{\text{ram}}/P_{\text{grav}}) > 1.2 \pm 0.2$ , tidal effects are important when  $\log_{10}(R_{\text{H1}}/r_t) > -0.5 \pm 0.1$  and satellite encounters are relevant for  $\log_{10}(E_s/E_{\text{int}}) > -1.25 \pm 0.25$ . By using these thresholds, we found that  $25 \pm 7$  per cent of the systems in our sample are perturbed by ram pressure,  $16 \pm 6$  per cent by tidal interactions with the host halo and  $10 \pm 3$  per cent by interactions with other satellites. For the latter, the typical perturber-to-system stellar mass ratio is 0.5. So, according to the simulation, the most common environmental mechanism that perturbs the H I in a galaxy with  $M_* > 10^9 M_\odot$  at redshift  $z = 0$  is ram pressure by the IGM. Considering that a system can be affected by more than one mechanism at the same time, we found that  $37 \pm 10$  per cent of the satellites in our sample are affected by at least one of the mechanisms proposed. This drops to  $9.1 \pm 2.5$  per cent if we also include in our sample central galaxies in the same range of stellar and H I masses considered, and assume that these systems are not affected by the environment. Note though that tidal interactions can perturb the H I of centrals, as can be seen in two rightmost panels in the central row of Fig. 10.

Fig. 11 shows how the satellites with perturbed H I are distributed in host halo mass  $M_{200}$  (top panel) and groupcentric distance  $d/r_{200}$  (bottom panel). Satellites that are affected by tidal interactions tend to be found in groups with  $M_{200} < 10^{14} M_\odot$  and are located within the virial radius of their group. Galaxies perturbed by ram pressure can be found also in more massive groups and at larger distances, although most (70 per cent) of them are located within the virial radius of their group. We have verified visually that those satellites considered to be perturbed by ram pressure and located at  $d > r_{200}$  indeed have disturbed H I morphologies. Finally, systems affected by interactions with other satellites exhibit a broader distribution both in  $d/r_{200}$  and in  $M_{200}$ , although they are more common in the most massive groups formed in the Ref-L100N1504 run. Given that satellite encounters are rare at  $z = 0$ , one may naively conclude that they have little impact on the evolution of the H I content of galaxies. We now show that this is not the case.

## 5.2 Environment and H I stripping

What are the mechanisms that deprive a satellite galaxy of its H I disc? We have seen that there is a fraction of satellites that experiences significant environmental perturbations at redshift  $z = 0$ . It is reasonable to expect that, for those satellites that are H I poor at  $z = 0$ , environmental mechanisms acted at some point in the past and contributed to the gas loss. To explore this scenario, we identified those satellites (with  $M_* > 10^9 M_\odot$ ) that are H I poor at  $z = 0$  in the Ref-L100N1504 run and followed their H I content, and the environmental processes that they experience, back to redshift  $z = 0.5$ .

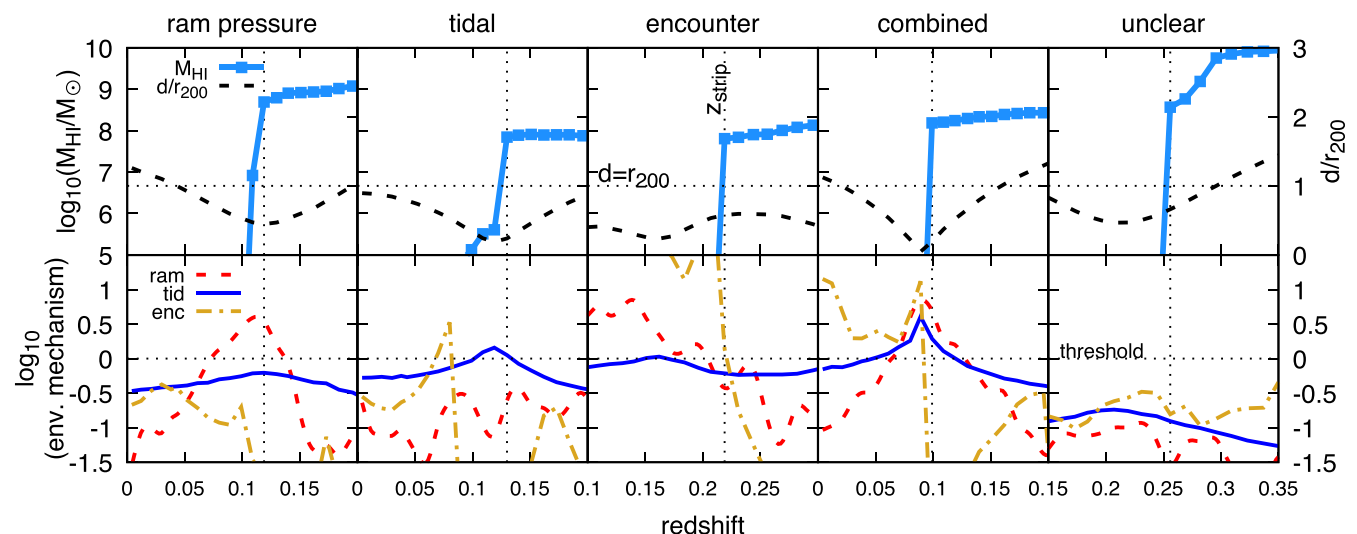
The tracking of each galaxy is achieved by identifying its main progenitor in the merger tree as described by Crain et al. (2016, see their appendix A). To achieve a good time resolution, we used specific EAGLE output files that are named ‘snipshots’. These outputs are saved with much higher frequency than the regular snapshots – approximately every 125 Myr for  $z < 0.5$  – and still contain all the



**Figure 11.** Distribution of  $M_{200}$  (top panel) and  $d/r_{200}$  (bottom panel) for the satellites in the Ref-L100N1504 run that are significantly perturbed by ram pressure by the IGM (red dashed line), tidal stripping by the host halo (blue solid line) or interactions with other satellites (yellow dot-dashed line) at redshift  $z = 0$ . Only satellites with  $M_* > 10^9 M_\odot$ ,  $M_{\text{H I}} > 6.8 \times 10^8 M_\odot$  are included. Galaxies perturbed by tidal interactions with the host halo are always located at  $d < r_{200}$  and prefer haloes with  $M_{200} < 10^{14}$ , while those perturbed by ram pressure or satellite interactions are more sparse and can be found in more massive groups.

data required to perform our calculations. Ram pressure and tidal interactions are computed as in Section 5.1, but here we use the stellar half-mass radius  $R_{\text{hm}}$  instead of  $R_{\text{H I}}$  as a reference radius. This because (a)  $R_{\text{H I}}$  is not defined for systems without H I, and (b) we are interested in the removal of the H I from the inner regions of the disc, thus  $R_{\text{hm}}$  provides a better motivated choice. In addition, at this radius we expect that the baryonic component dominates the gravitational restoring acceleration, thus the right-hand side of equation (2) can be more conveniently computed as  $2\pi G \Sigma_{\text{bar}} \Sigma_{\text{ISM}}$  (Gunn & Gott 1972), where  $G$  is the gravitational constant, and  $\Sigma_{\text{ISM}}$  and  $\Sigma_{\text{bar}}$  are the ISM and the stellar+ISM surface densities at  $R = R_{\text{hm}}$ . To deal with gas-depleted systems for which  $\Sigma_{\text{ISM}} = 0$ , we assumed a lower limit of  $1 M_\odot \text{pc}^{-2}$  for the ISM surface density.

In total, we track 2482 H I-poor satellites from  $z = 0$  to 0.5, recording at each time step their stellar and H I masses, the  $M_{200}$  of their current group, their distance to the centre of the group and the magnitudes of three environmental effects. For each of these systems, we determined the redshift at which the H I is stripped from the galaxy,  $z_{\text{strip}}$ , as the snapshot at which the H I mass crosses a threshold of  $10^7 M_\odot$ , and stays above it in the two snapshots at higher  $z$ . We found that the H I mass of satellites is either zero or well above this threshold (see top panels in Fig. 12), which justifies our choice. We verified that using a different threshold has a negligible impact on our results. We found that 1613 (65 per cent) of these satellites have lost their H I in this redshift range. For these systems, we attempted to identify which environmental effect is contributing to the gas stripping. Unfortunately this procedure is not straightforward given that our environment estimators are crude proxies for the true ongoing environmental processes. In addition, it is often the case that more than one mechanism acts at the same time, which complicates further the analysis. The approach that we adopted is to carefully define ad hoc thresholds above which a given process is significant, similarly to that done in Section 5.1. To achieve this, we first produced plots of  $P_{\text{ram}}/P_{\text{grav}}$ ,  $R_{\text{hm}}/r_t$  and  $E_s/E_{\text{int}}$  as a function of redshift for each of the 1613 gas-depleted



**Figure 12.** H I masses and the magnitude of environmental effects as a function of redshift for five H I-poor satellites extracted from the Ref-L100N1504 EAGLE run at redshift  $z = 0$ . Each case is representative for H I stripping due to a different environmental process, as indicated above each column. The top panels show the evolution of  $M_{\text{H I}}$  (symbols connected with solid lines, left-hand axis) and of  $d/r_{200}$  (dashed line, right-hand axis). The bottom panels show the evolution of the environmental mechanisms: ram pressure from the IGM ( $\log_{10}(P_{\text{ram}}/P_{\text{grav}})$ , dashed line), tidal interaction with the host halo ( $\log_{10}(R_{\text{hm}}/r_t)$ , solid line) and satellite encounters ( $\log_{10}(E_s/E_{\text{int}})$ , dot-dashed line). To emphasize the mechanisms that contribute to the stripping, all values have been normalized by their adopted threshold values (horizontal dotted lines in the bottom panels, see text). The vertical dotted lines show the redshift at which the H I is stripped.

satellites. Then, by inspecting these plots, we identified those cases where a given mechanism is clearly the dominant stripping process.

For the case of ram pressure (or tidal) stripping, we focus on those systems where the removal of gas is associated with a clear peak in  $P_{\text{ram}}/P_{\text{grav}}$  (or  $R_{\text{hm}}/r_t$ ), whereas the other estimators do not show peculiar features around  $z_{\text{strip}}$  and/or are well below their assumed thresholds. In this way we determine that ram pressure contributes to the stripping when  $\log_{10}(P_{\text{ram}}/P_{\text{grav}}) > 1.3 \pm 0.2$ , while tidal stripping occurs for  $\log_{10}(R_{\text{hm}}/r_t) > -0.7 \pm 0.1$ . The case of satellite encounters is more complex. In most galaxies, we found that the gas removal is associated with a clear discontinuity in  $E_s/E_{\text{int}}$ , which typically increases by more than one order of magnitude in the two snippets before and after  $z_{\text{strip}}$ . Given that  $E_s/E_{\text{int}}$  varies smoothly before and after the discontinuity at  $z = z_{\text{strip}}$ , there is little doubt that we can consider it as a signature of gas removal due to a high-speed encounter. Thus we consider satellite encounters relevant to gas removal when  $\Delta \log_{10}(E_s/E_{\text{int}}) > 0.7 \pm 0.2$ , where the difference  $\Delta$  is computed around  $z_{\text{strip}}$ . We stress that variations in energy ratio with redshift are usually well below 0.5 dex, and clear discontinuities can be seen only during these high-speed encounters. Additionally, we require the energy ratio after the removal to be larger than a given threshold. This is the major source of uncertainty in our calculation: based on our previous findings (Section 5.1) we set this threshold at  $\log_{10}(E_s/E_{\text{int}}) > -1.5 \pm 0.5$ , as in this range perturbations in the H I morphology should arise. We stress that the error bars reported for all these thresholds have no strict statistical meaning, but they have been chosen ad hoc as representative for the uncertainties intrinsic in our procedure.

Fig. 12 shows five representative cases where the H I loss appears to be caused by a distinctive mechanism. The upper panels show how the H I mass and the groupcentric distance of the system vary with the redshift, the bottom panels show (on a logarithmic scale) the various environmental effects which have been re-scaled to their respective thresholds (horizontal dotted lines). In the first two cases, the stripping happens when  $d/r_{200}$  reaches a local minimum, i.e. at the pericentre of the satellite's orbit. In the first example, the ram pressure peaks and dominates over the other environmental effects at  $z = z_{\text{strip}}$ , tidal forces are significant but not larger than the adopted threshold. The opposite happens in the second example, where tidal forces dominates while the other mechanisms are not relevant and do not show any peculiar feature at the moment of the stripping. In the third case, the stripping happens *before* the satellite reaches the pericentre because of a powerful energy injection due to a satellite encounter, which changes the system's internal energy by several orders of magnitude. The next example shows a situation where more than one mechanism contributes to the gas removal: in this particular case, all processes are simultaneously at work. Note that, in the last two examples,  $E_s/E_{\text{int}}$  remains large after the stripping event. This is not due to further pair interactions, which would feature as additional discontinuities in  $E_s/E_{\text{int}}$ , but rather to the decline (in modulus) of the internal energy of the system after the encounter. Finally, we classify the last case as 'unclear', because all mechanisms are below the threshold and no peculiar features can be seen around  $z = z_{\text{strip}}$ . Note also that in all these examples the gas removal occurs at  $d < r_{200}$ .

Situations where ram pressure or tidal interactions alone are responsible for the gas removal are very infrequent, as they occur only in a small fraction of cases. It is instead more common that a combination of satellite encounters and ram pressure by the IGM are the cause of the gas loss. Regardless of whether or not these mechanisms work alone or together, we found that ram pressure,

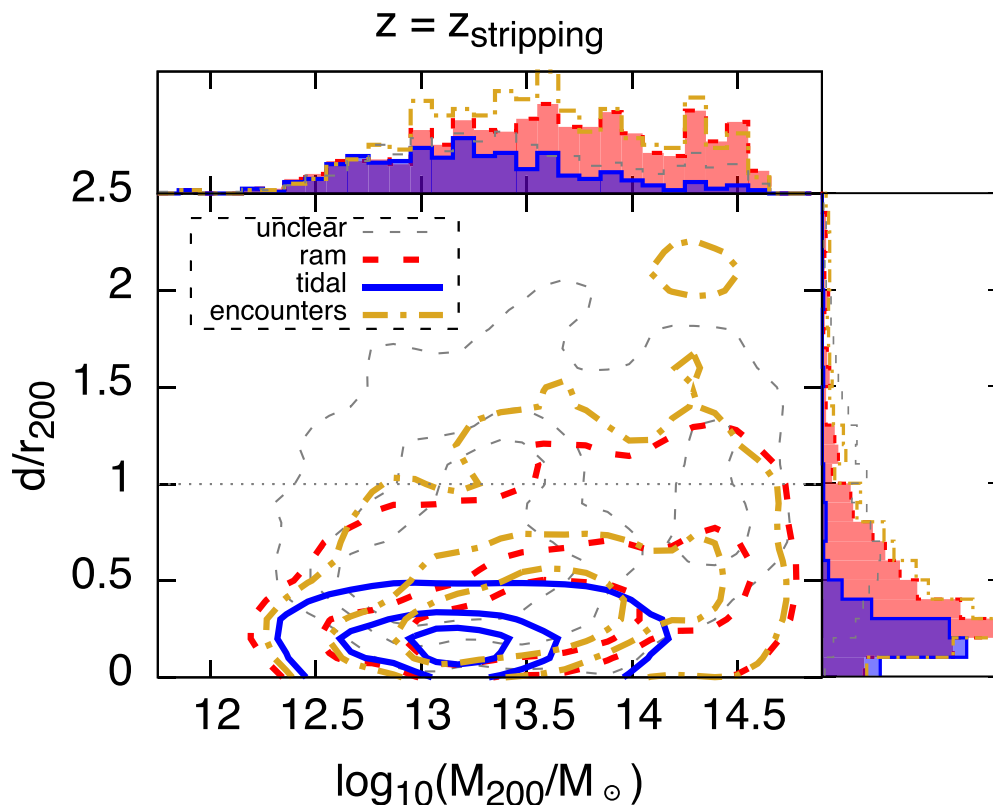
tidal interactions and satellite encounters are relevant to the stripping in, respectively,  $47 \pm 8$ ,  $21 \pm 7$  and  $55 \pm 11$  percent of the cases analysed. Thus, it seems that the third of the environmental mechanisms proposed is the most common mode by which currently H I-poor satellite galaxies with  $M_* > 10^9 M_{\odot}$  have lost their H I content in the past. This finding might appear to contradict the results of Section 5.1, where we showed that satellite encounters are the least common mechanism that perturbs H I in galaxies at redshift zero. We believe that a simple explanation is that different environmental mechanisms act on different time-scales: at a given time, the probability of observing a high-speed encounter between satellites is smaller than that of observing ongoing ram pressure stripping by the IGM. A mechanism that acts on short time-scales would also explain why the environment acts mainly as an on/off switch for the H I content of satellites, as we found in Section 4.1.

For each of the 1613 satellites, we estimate the duration of the stripping process as follows. We set the beginning of the H I stripping to the minimum redshift above  $z_{\text{strip}}$  at which the fractional variation in H I mass ( $\Delta M_{\text{H I}}/M_{\text{H I}}$ ) computed between two consecutive snippets is below a threshold of 0.1. The end of the H I stripping is set to the maximum redshift below  $z_{\text{strip}}$  at which the H I mass falls below  $1.36 \times 10^6 M_{\odot}$  (i.e. the hydrogen mass of a single gas particle). H I masses are linearly interpolated between consecutive snippets in order to refine the calculation. We find that the distribution of the stripping duration has median  $t_{\text{strip}} = 230$  Myr, and 16th and 84th percentiles of 100 and 518 Myr.  $t_{\text{strip}}$  is not very sensitive to the exact choice of the threshold in  $\Delta M_{\text{H I}}/M_{\text{H I}}$ , as using  $\Delta M_{\text{H I}}/M_{\text{H I}} = 0.01$  (0.3) yields  $t_{\text{strip}} = 280$  Myr (200 Myr). Also, it depends weakly on  $M_*$ ,  $M_{200}$  and on the environmental mechanism considered: for instance,  $t_{\text{strip}} = 210$  Myr (245 Myr) for stripping clearly dominated by satellite encounters (ram pressure). However, our analysis is limited by the fact that  $t_{\text{strip}}$  is of the same order of the snippet time step ( $\sim 125$  Myr), which precludes the sampling of stripping events shorter than this value. Thus it is possible that the typical time-scale for satellite encounters is much shorter than that for ram pressure stripping, as suggested above. In general, we can robustly conclude that the time-scale for H I removal by direct environmental processes is typically  $< 500$  Myr. Interestingly, this is comparable to the quenching time-scale for satellite galaxies in clusters inferred by Muzzin et al. (2014). As discussed by Trayford et al. (2016), such a short quenching time-scale would lead to a 'green valley' with properties consistent with those observed.

We point out that, in those cases where the H I removal is solely due to satellite-satellite interactions, it is still possible that what ultimately causes the system to lose its gas is the ram pressure from the IGM. In fact, the gravitational restoring force that we compute refers to the case where the stars and the gas settle on to the midplane in a regular disc, and may be overestimated if the gas is misplaced from this configuration. Thus,  $P_{\text{ram}}/P_{\text{grav}}$  may be underestimated during the interaction with satellites. Note that we exclude from the computation of  $P_{\text{ram}}$  the gas particles that are gravitationally bound to the other satellites, which may in principle contribute to the ram pressure during the satellite encounters. We verified however that including these particles increases the fraction of ram pressure stripped occurrences only by a few per cent.

Interestingly, it is believed that the Magellanic Stream – the spectacular H I trail that the Magellanic Clouds are leaving behind as they move around the Milky Way – is produced by the mutual interactions between the clouds, and possibly amplified by the ram pressure by the Galaxy's hot halo (Besla et al. 2012; Salem et al. 2015). Despite the uncertainties of our analysis, our conclusions are consistent with those of Besla et al. (2012) and Pearson et al. (2016),





**Figure 13.** Distribution in the plane  $(M_{200}, d/r_{200})$  for the satellites that are H I poor at redshift  $z = 0$  evaluated at the moment of their H I loss in the Ref-L100N1504 run. Only satellites with  $M_*(z = 0) > 10^9 M_\odot$  are considered. Different environmental mechanisms are shown: tidal stripping by the host-halo (solid lines), ram pressure stripping by the IGM (thick dashed lines), satellite–satellite interactions (dot–dashed lines) or none of the above (thin dashed lines). Each mechanism is shown by three contours, representing (from the innermost to the outermost) the 30, 60 and 90 per cent of the enclosed systems. The horizontal dotted line is at  $d = r_{200}$ . The histograms on the top and on the right show the marginalized distributions. Tidal stripping acts within  $0.5 r_{200}$  in haloes with  $M_{200} < 10^{14} M_\odot$ , the other mechanisms work at larger distances and in more massive haloes as well.

i.e. that satellite–satellite interactions play a key role in shaping the evolution of galaxies in the stellar mass range considered.

In Fig. 13 we show how the 1613 gas-depleted satellites of our sample populate the  $(M_{200}, d/r_{200})$  plane at the time of their H I loss. Different colours are used to separate the different environmental mechanisms, and the contours enclose 30, 60 and 90 per cent of the systems. Clearly, tidal stripping operates nearly always (95 per cent of the cases) within *half* the virial radius of the group, and rarely occurs in host haloes more massive than  $\sim 10^{14} M_\odot$ . Stripping due to ram pressure and satellite interactions, instead, is common at all halo masses and can occur at a larger distances from the group centre, but usually inside the virial radius (in 89 and 82 per cent of the cases, respectively). Our results are in agreement with those of Bahé et al. (2013) who found that direct environmental stripping of cold gas occurs preferentially inside  $r_{200}$ , whereas stripping of the hot halo can occur at distances up to  $5 r_{200}$ . Note that ram pressure and satellite interactions operate in a similar region of the  $(M_{200}, d/r_{200})$  space, suggesting that an interplay between these two processes is possible, as discussed above. To gain insights into the origin of this interplay, we compared the average IGM density radial profile with the average subhalo number density radial profile derived by stacking groups of similar  $M_{200}$  in Ref-L100N1504. The IGM density profiles are evaluated by excluding all gas particles bound to satellites. We found that, for  $10^{12.5} < M_{200} < 10^{14} M_\odot$ , the two profiles have a very similar shape within  $r_{200}$ , indicating the existence of a strong correlation between the intra-

group gas density and the subhalo number density in this halo mass range.

In Fig. 13 we also report those satellites whose H I loss cannot be unambiguously related to one of the environmental mechanisms considered (thin dashed line). These galaxies constitute a significant fraction ( $31 \pm 6$  per cent) of our sample, and are also those located farther from the group centres: 60 per cent of them are within  $r_{200}$ , only 22 per cent within  $0.5 r_{200}$ . Also, they are more frequent at larger  $M_*$ , and dominate our satellite sample at  $M_* > 10^{10} M_\odot$ . One might argue that in these systems the H I removal is due to a combination of environmental processes acting together, each of which individually would not have been strong enough to strip the gas. However, in many of these ‘unclear’ cases, none of the direct environmental processes considered shows significant enhancement around  $z = z_{\text{strip}}$ . Although we can speculate that both internal processes (e.g. AGN or stellar feedback) and indirect environmental processes (e.g. starvation) can contribute to the gas loss in these systems, a detailed analysis of these mechanisms is beyond the scope of this study. Our findings simply indicate that it is unlikely that in these systems the H I loss was caused solely by one of the environmental mechanisms proposed.

## 6 CONCLUSIONS

In this work we used the EAGLE suite of cosmological simulations to study how the H I content of galaxies is affected by their

environment, for which we used the host halo virial mass  $M_{200}$  as a proxy. We first compared the predictions of *EAGLE* and of three SA models of galaxy evolution to the observed H I-environment trends found by *Fab12* and *Cat13*. Then, we investigated how the H I content of satellite galaxies with stellar mass  $M_* > 10^9 M_\odot$  changes as a function of  $M_{200}$  and  $M_*$ . Finally, we discussed which environmental processes directly affect the H I content of a satellite and eventually remove it from its disc during its evolution.

Our findings can be summarized as follows.

(i) Galaxies in *EAGLE* follow the observed H I-environment trends found by *Fab12* and *Cat13* remarkably well, whereas the SA models predict systems that are too H I rich in dense environments (see Figs 1 and 4). A possible cause of this discrepancy is that SA models lack a self-consistent treatment of direct removal of galaxies' cold ISM.

(ii) By focusing on satellites, we found that the main effect of the environment is to control whether or not galaxies have H I at all, rather than producing a continuous trend. At a fixed  $M_*$ , the fraction of systems nearly devoid of H I increases with increasing host halo mass  $M_{200}$  as a result of stronger environmental effects, and at a fixed  $M_{200}$  it decreases with increasing  $M_*$  as the gas is confined by deeper potential wells (see Fig. 6).

(iii) The H I deficiency increases within 1–2 virial radii from the group centre in haloes with  $M_{200} > 10^{13} M_\odot$  (see Fig. 8), as also revealed by observations (Solanes et al. 2001; Gavazzi et al. 2006).

(iv) Satellites devoid of H I tend to be concentrated within the virial radius of their group, whereas H I-rich satellites are more sparsely distributed (see Fig. 9).

(v) At redshift  $z = 0$ , the most common environmental mechanism that can visibly affect the H I morphology of a satellite with  $M_{H I} > 6.8 \times 10^8 M_\odot$  is the ram pressure of the IGM, while tidal interactions with the host halo and satellite–satellite encounters are less frequent. Tidal interactions are confined to the virial radius of haloes with  $M_{200} < 10^{14} M_\odot$ , ram pressure and satellite encounters can be effective also at larger distances and in more massive haloes (see Fig. 11).

(vi) By tracking back in time the H I content and the environmental properties of satellites that are devoid of H I at redshift  $z = 0$ , we found that the most common stripping mechanism is satellite–satellite interactions, followed by ram pressure and tidal stripping. The time-scale for H I removal is typically less than 0.5 Gyr. Tidal stripping occurs nearly always within  $0.5 r_{200}$  in haloes with  $M_{200} < 10^{14} M_\odot$ , the other mechanisms act also in more massive haloes, usually within  $r_{200}$  (see Fig. 13). We suggest that satellite–satellite interactions act on much shorter time-scales than the other processes, which would explain their relative rarity at any particular redshift. Finally, in about a third of the cases, the H I stripping could not be unambiguously related to any of the environmental mechanisms analysed.

The next generation of H I surveys will have the capabilities to corroborate or reject many of the predictions advanced in this study. We stress that the analysis presented here was based on galaxies with  $M_* > 10^9 M_\odot$ , for which the observed and simulated stellar mass functions are in good agreement with each other. The study of the Local Group, however, indicates that the role of the environment is crucial to explain the properties of the faintest galaxies observed. In the future, it will be interesting to extend the analysis presented in this work to the dwarf and ultrafaint dwarf galaxy regimes. The *APOSTLE* suite of cosmological zoom-in simulations of the Local Group environment (Sawala et al. 2016; Fattahi et al. 2016) is well suited for this purpose.

## ACKNOWLEDGEMENTS

AM kindly thanks B. Catinella and S. Fabello for providing their results, and P. Serra for useful discussions. AM and TvdH acknowledge support from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement no. 291531 and 278594-GasAroundGalaxies. RAC is a Royal Society University Research Fellow. This work used the DiRAC Data Centric system at Durham University, operated by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility ([www.dirac.ac.uk](http://www.dirac.ac.uk)). This equipment was funded by BIS National E-infrastructure capital grant ST/K00042X/1, STFC capital grant ST/H008519/1 and STFC DiRAC Operations grant ST/K003267/1 and Durham University. DiRAC is part of the National E-Infrastructure.

## REFERENCES

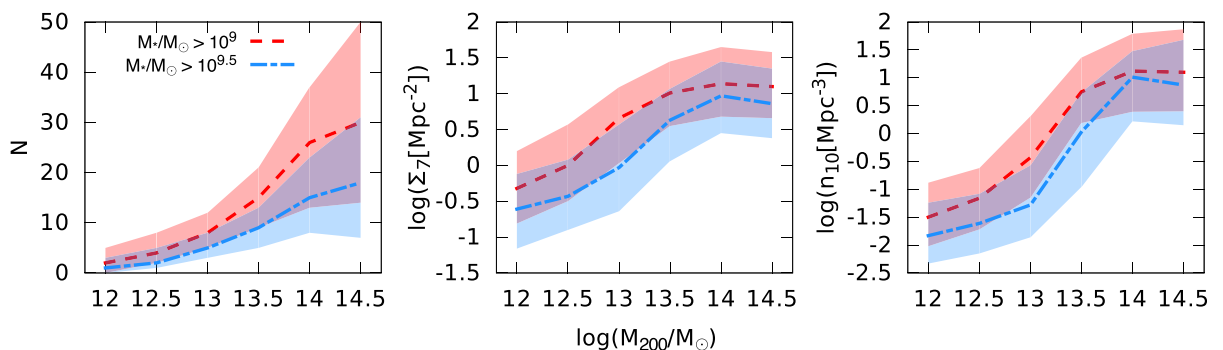
- Agertz O. et al., 2007, *MNRAS*, 380, 963  
 Bahé Y. M., McCarthy I. G., Balogh M. L., Font A. S., 2013, *MNRAS*, 430, 3017  
 Bahé Y. M. et al., 2016, *MNRAS*, 456, 1115  
 Baldry I. K., Glazebrook K., Brinkmann J., Ivezić Ž., Lupton R. H., Nichol R. C., Szalay A. S., 2004, *ApJ*, 600, 681  
 Balogh M. L., Baldry I. K., Nichol R., Miller C., Bower R., Glazebrook K., 2004, *ApJ*, 615, L101  
 Bekki K., Couch W. J., Shioya Y., 2002, *ApJ*, 577, 651  
 Besla G., Kallivayalil N., Hernquist L., van der Marel R. P., Cox T. J., Kereš D., 2012, *MNRAS*, 421, 2109  
 Binney J., Tremaine S., 2008, *Galactic Dynamics*, 2nd edn. Princeton Univ. Press, Princeton, NJ  
 Blitz L., Rosolowsky E., 2006, *ApJ*, 650, 933  
 Booth C. M., Schaye J., 2009, *MNRAS*, 398, 53  
 Boylan-Kolchin M., Springel V., White S. D. M., Jenkins A., Lemson G., 2009, *MNRAS*, 398, 1150  
 Broeils A. H., Rhee M.-H., 1997, *A&A*, 324, 877  
 Catinella B. et al., 2010, *MNRAS*, 403, 683  
 Catinella B. et al., 2013, *MNRAS*, 436, 34 (*Cat13*)  
 Chabrier G., 2003, *PASP*, 115, 763  
 Chung A., van Gorkom J. H., Kenney J. D. P., Crowl H., Vollmer B., 2009, *AJ*, 138, 1741  
 Cooper M. C. et al., 2006, *MNRAS*, 370, 198  
 Cortese L., Catinella B., Boissier S., Boselli A., Heinis S., 2011, *MNRAS*, 415, 1797  
 Crain R. A. et al., 2009, *MNRAS*, 399, 1773  
 Crain R. A. et al., 2015, *MNRAS*, 450, 1937  
 Crain R. A. et al., 2016, preprint ([arXiv:e-prints](https://arxiv.org/abs/1608.07648))  
 Cullen L., Dehnen W., 2010, *MNRAS*, 408, 669  
 Dalla Vecchia C., Schaye J., 2012, *MNRAS*, 426, 140  
 Dolag K., Borgani S., Murante G., Springel V., 2009, *MNRAS*, 399, 497  
 Dressler A., 1980, *ApJ*, 236, 351  
 Durier F., Dalla Vecchia C., 2012, *MNRAS*, 419, 465  
 Fabello S., Kauffmann G., Catinella B., Li C., Giovanelli R., Haynes M. P., 2012, *MNRAS*, 427, 2841 (*Fab12*)  
 Fattahi A. et al., 2016, *MNRAS*, 457, 844  
 Few C. G., Gibson B. K., Courty S., Michel-Dansac L., Brook C. B., Stinson G. S., 2012, *A&A*, 547, A63  
 Furlong M. et al., 2015, *MNRAS*, 450, 4486  
 Gavazzi G., O'Neil K., Boselli A., van Driel W., 2006, *A&A*, 449, 929  
 Giovanelli R., Haynes M. P., 1985, *ApJ*, 292, 404  
 Giovanelli R. et al., 2005, *AJ*, 130, 2598  
 Grcevich J., Putman M. E., 2009, *ApJ*, 696, 385  
 Gunn J. E., Gott J. R., III, 1972, *ApJ*, 176, 1  
 Guo Q. et al., 2011, *MNRAS*, 413, 101 (*Guo11*)  
 Guo Q., White S., Angulo R. E., Henriques B., Lemson G., Boylan-Kolchin M., Thomas P., Short C., 2013, *MNRAS*, 428, 1351 (*Guo13*)

- Haardt F., Madau P., 2001, in Neumann D. M., Tran J. T. V., eds, Clusters of Galaxies and the High Redshift Universe Observed in X-rays. CEA, Saclay, p. 64
- Haas M. R., Schaye J., Jeon-Daniel A., 2012, MNRAS, 419, 2133
- Haynes M. P., Giovanelli R., 1984, AJ, 89, 758
- Haynes M. P., Giovanelli R., 1986, ApJ, 306, 466
- Haynes M. P. et al., 2011, AJ, 142, 170
- Henriques B. M. B., White S. D. M., Thomas P. A., Angulo R., Guo Q., Lemson G., Springel V., Overzier R., 2015, MNRAS, 451, 2663 (Hen15)
- Hess K. M., Wilcots E. M., 2013, AJ, 146, 124
- Hu C.-Y., Naab T., Walch S., Moster B. P., Oser L., 2014, MNRAS, 443, 1173
- Jaffé Y. L., Smith R., Candlish G. N., Poggianti B. M., Sheen Y.-K., Verheijen M. A. W., 2015, MNRAS, 448, 1715
- Kauffmann G. et al., 2003, MNRAS, 341, 54
- Kauffmann G., White S. D. M., Heckman T. M., Ménard B., Brinchmann J., Charlot S., Tremonti C., Brinkmann J., 2004, MNRAS, 353, 713
- Kaufmann T., Mayer L., Wadsley J., Stadel J., Moore B., 2006, MNRAS, 370, 1612
- Kennicutt R. C., Jr, 1998, ApJ, 498, 541
- Komatsu E. et al., 2011, ApJS, 192, 18
- Lagos C. D. P., Baugh C. M., Lacey C. G., Benson A. J., Kim H.-S., Power C., 2011, MNRAS, 418, 1649
- Lagos C. d. P. et al., 2015, MNRAS, 452, 3815
- Lagos C. d. P. et al., 2016, MNRAS, 459, 2632
- Limousin M., Sommer-Larsen J., Natarajan P., Milvang-Jensen B., 2009, ApJ, 696, 1771
- Muzzin A. et al., 2014, ApJ, 796, 65
- Oemler A., Jr, 1974, ApJ, 194, 1
- Oosterloo T., van Gorkom J., 2005, A&A, 437, L19
- Pawlik A. H., Schaye J., 2008, MNRAS, 389, 651
- Pearson S. et al., 2016, MNRAS, 459, 1827
- Planck Collaboration XVI, 2014, A&A, 571, A16
- Price D. J., 2008, J. Comput. Phys., 227, 10040
- Rafieferantsoa M., Davé R., Anglés-Alcázar D., Katz N., Kollmeier J. A., Oppenheimer B. D., 2015, MNRAS, 453, 3980
- Rahmati A., Pawlik A. H., Raicevic M., Schaye J., 2013, MNRAS, 430, 2427
- Rahmati A., Schaye J., Crain R. A., Oppenheimer B. D., Schaller M., Theuns T., 2016, MNRAS, 459, 310
- Roediger E., Brüggen M., 2007, MNRAS, 380, 1399
- Rosas-Guevara Y. M. et al., 2015, MNRAS, 454, 1038
- Salem M., Besla G., Bryan G., Putman M., van der Marel R. P., Tonnesen S., 2015, ApJ, 815, 77
- Sawala T. et al., 2016, MNRAS, 457, 1931
- Schaller M., Dalla Vecchia C., Schaye J., Bower R. G., Theuns T., Crain R. A., Furlong M., McCarthy I. G., 2015, MNRAS, 454, 2277
- Schaye J., Dalla Vecchia C., 2008, MNRAS, 383, 1210
- Schaye J. et al., 2015, MNRAS, 446, 521
- Solanes J. M., Manrique A., García-Gómez C., González-Casado G., Giovanelli R., Haynes M. P., 2001, ApJ, 548, 97
- Spekkens K., Urbancic N., Mason B. S., Willman B., Aguirre J. E., 2014, ApJ, 795, L5
- Spergel D. N. et al., 2003, ApJS, 148, 175
- Springel V., 2005, MNRAS, 364, 1105
- Springel V. et al., 2005a, Nature, 435, 629
- Springel V., Di Matteo T., Hernquist L., 2005b, MNRAS, 361, 776
- Tonnesen S., Bryan G. L., van Gorkom J. H., 2007, ApJ, 671, 1434
- Trayford J. W. et al., 2015, MNRAS, 452, 2879
- Trayford J. W., Theuns T., Bower R. G., Crain R. A., Lagos C. d. P., Schaller M., Schaye J., 2016, MNRAS, in press
- Villalobos Á., De Lucia G., Borgani S., Murante G., 2012, MNRAS, 424, 2401
- Wetzel A. R., Tinker J. L., Conroy C., 2012, MNRAS, 424, 232
- Wiersma R. P. C., Schaye J., Smith B. D., 2009a, MNRAS, 393, 99
- Wiersma R. P. C., Schaye J., Theuns T., Dalla Vecchia C., Tornatore L., 2009b, MNRAS, 399, 574
- Yang X., Mo H. J., van den Bosch F. C., Zhang Y., Han J., 2012, ApJ, 752, 41
- York D. G. et al., 2000, AJ, 120, 1579
- Yun M. S., Ho P. T. P., Lo K. Y., 1994, Nature, 372, 530

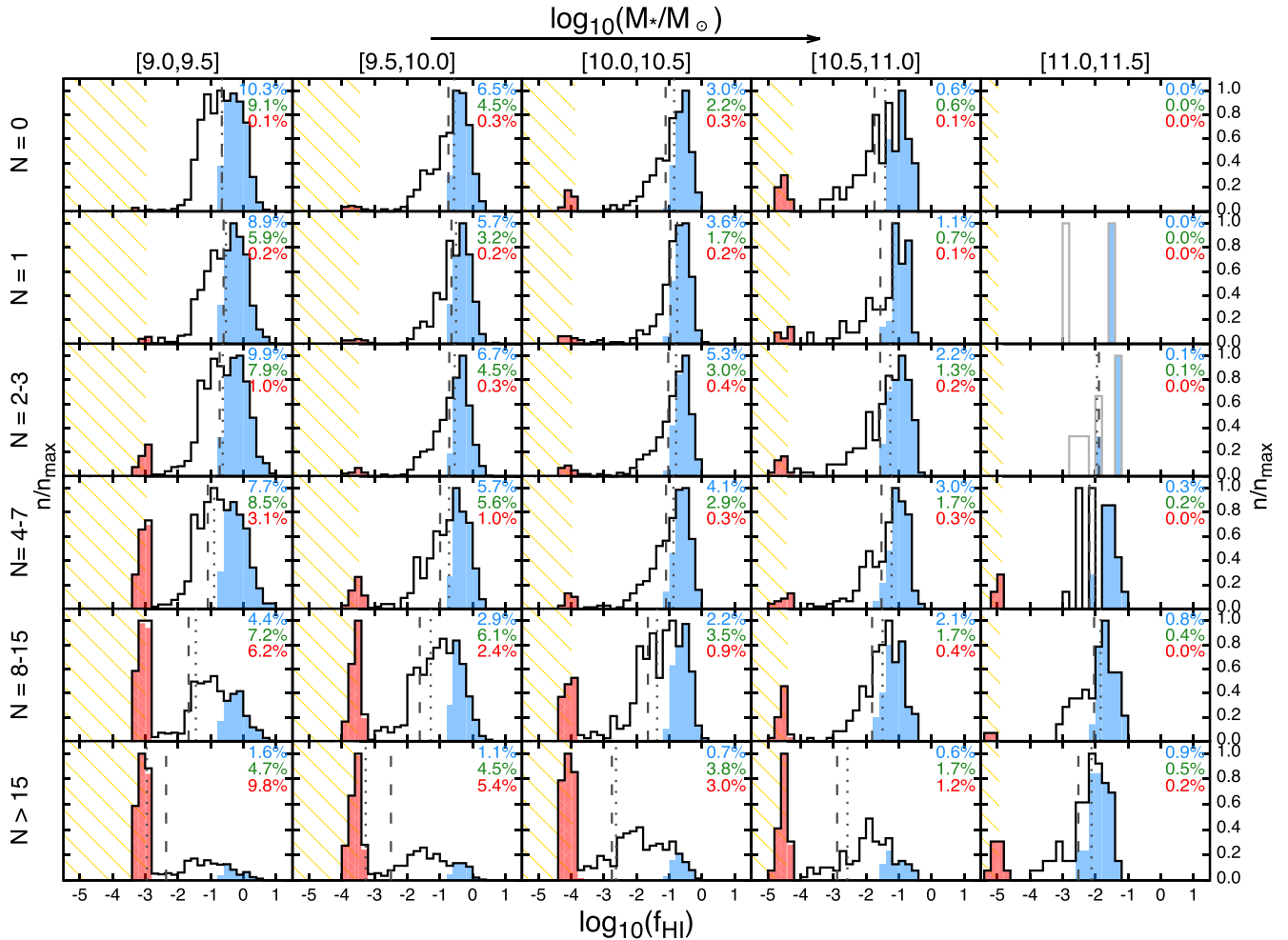
## APPENDIX A: COMPARING DIFFERENT ENVIRONMENT ESTIMATORS

The definition of environment adopted in this work is based on quantities that describe the local matter density. In cosmological simulations, a commonly adopted way to define the environment of a galaxy is the virial mass  $M_{200}$  of the FoF group to which that galaxy belongs. This quantity is difficult to measure in real observations, where the environment is often evaluated in terms of galaxy counts or number density. One may wonder how these quantities relate to  $M_{200}$ .

We used the EAGLE simulations to investigate the correlation between  $M_{200}$  and three commonly used ‘observable’ environment estimators: the number  $N$  of neighbours within a fixed 2D aperture of 1 Mpc and  $\pm 500 \text{ km s}^{-1}$  in line-of-sight velocity (left-hand panel in Fig. A1), the surface density of galaxies  $\Sigma_7$  up to the seventh nearest (in projected distance) neighbour within  $\pm 500 \text{ km s}^{-1}$  in line-of-sight velocity (central panel in Fig. A1) and the volume density of galaxies  $n_{10}$  up to the 10th nearest (in 3D distance) neighbour (right-hand panel in Fig. A1). These quantities are defined by



**Figure A1.** Correlation between  $M_{200}$  and three different environment estimators based on galaxy density ( $N$  in the left-hand panel,  $\Sigma_7$  in the central panel,  $n_{10}$  in the right-hand panel, see text for the details) in the Ref-L100N1504 run. Each environment estimator is determined twice, using galaxies above stellar masses of either  $10^9 M_\odot$  (red) or  $10^{9.5} M_\odot$  (blue). Lines represent the median values, while shaded regions bracket the 16th and 84th percentiles. All environment estimators correlate with  $M_{200}$ , with some scatter.



**Figure A2.** As for Fig. 6, but here (a) the environment is characterized in terms of  $N$ , the number of galaxies with  $M_* > 10^9 M_\odot$  within a fixed 2D aperture of 1 Mpc and  $\pm 500 \text{ km s}^{-1}$  in line-of-sight velocity; (b) the  $\text{H I}$  masses are computed in circular apertures of 150 kpc in diameter and line-of-sight velocity ranges of  $\pm 400 \text{ km s}^{-1}$ , by analogy with ALFALFA observations; (c) all galaxies, and not only the satellites, are considered. The trends in  $M_*$  and  $M_{200}$  are similar to those shown in Fig. 6, but here the fraction of  $\text{H I}$ -poor galaxies are reduced.

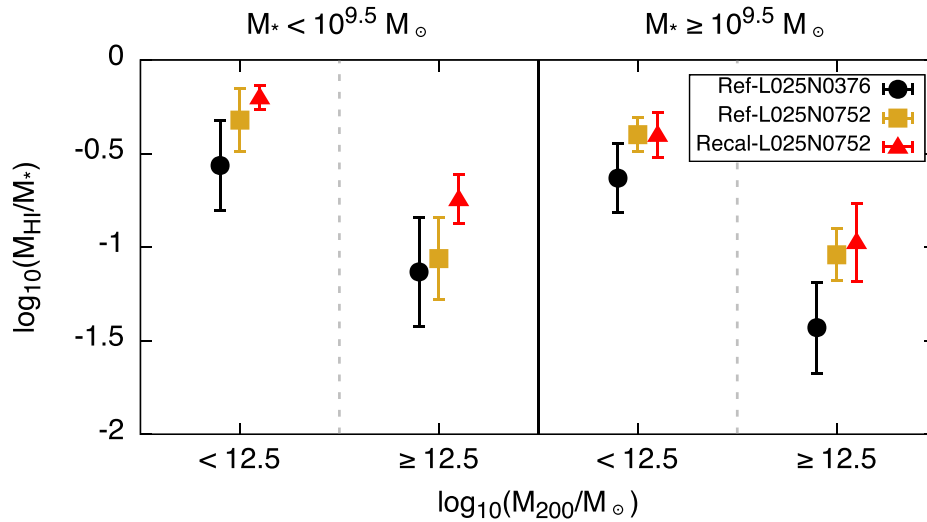
using neighbours with stellar masses (or absolute magnitude) above a given threshold, thus we show separately two cases adopting either  $M_* > 10^9$  or  $> 10^{9.5} M_\odot$ . Lines in Fig. A1 show the median of the distribution at a given  $M_{200}$ , the shaded region brackets the 16th and 84th percentiles. In all cases, there is a monotonic relation between  $M_{200}$  and the other environment estimators, in agreement with the analysis of a SA model by Haas et al. (2012). However, the scatter in these relations is large: for instance, a value of 0.5 in  $\log_{10}(\Sigma_7)$  may correspond to anything between  $10^{12.5}$  and  $10^{14.5} M_\odot$  in virial mass.

It would be interesting to verify whether or not the environmental trends shown in Fig. 6 are maintained if (a) we use one of the observable proxies for the environment rather than the host halo mass; (b) we compute the  $\text{H I}$  masses as in ALFALFA observations (see Sections 2.1 and 3.1) and (c) we include centrals in our analysis. In Fig. A2 we show the distribution of  $\log_{10} M_{\text{H I}}/M_*$  for all *EAGLE* galaxies with  $M_* > 10^9 M_\odot$  in different bins of stellar mass and environment density  $N$ . The bins in  $N$  have been arranged in order to have an approximately equal number of galaxies in each bin,

regardless of the stellar mass. In general, the environmental trends shown in Fig. A2 are similar to those presented by using  $M_{200}$  as an environment proxy: at a fixed  $M_*$ ,  $\text{H I}$ -poor galaxies become dominant at large  $N$ , while their number decreases with increasing  $M_*$  at a fixed  $N$  (although this is visible mainly for  $N > 4$ ). The main difference however is that the fraction of  $\text{H I}$ -poor systems is severely reduced with respect to Fig. 6. This is partially due to the inclusion of centrals, which are virtually never  $\text{H I}$  poor. It is also a consequence of the method adopted to compute the  $\text{H I}$  masses, which may now be overestimated given the aperture and velocity range used.

Finally, we stress that  $M_{200}$  does not necessarily provides an optimal definition for the local matter density. Consider a virialized massive group or a cluster of galaxies, for instance. All galaxies in this system share the same host halo mass. However, satellites at the centre of the system experience more ram pressure stripping and tidal interactions than those that reside in the outskirts. Thus, a definition of the environment based on galaxy number counts might be preferable.





**Figure B1.** Numerical convergence test. Median  $f_{\text{H I}}$  for satellites that reside in haloes less or more massive than  $10^{12.5} M_{\odot}$  in the Ref-L025N0376 (circles), Ref-L025N0752 (squares) and Recal-L025N0752 (triangles) runs. The left-hand panel shows satellites with  $M_{*} < 10^{9.5} M_{\odot}$ , the right-hand panel shows more massive systems. A minimum  $\text{H I}$  mass of  $1.36 \times 10^6 M_{\odot}$ , corresponding to the  $\text{H I}$  mass resolution of the Ref-L025N0376 run, is assumed. Error bars represent the  $1\sigma$  uncertainty on the median and are derived by bootstrap resampling the galaxies in each bin.

## APPENDIX B: STRONG AND WEAK CONVERGENCE TESTS

In this section we test the numerical convergence of our results. Schaye et al. (2015) introduced the concepts of ‘strong’ and ‘weak’ convergence tests. The former refers to the case where a simulation is re-run at a different resolution by adopting the same subgrid physics and parameters as the original ‘fiducial’ run, while the latter refers to the case where the subgrid physics and parameters are re-calibrated to recover similar agreement with the chosen calibration diagnostics (which in *EAGLE* are the galaxy stellar mass function and the mass–size relation at  $z = 0.1$ ). The *EAGLE* suite has two high-resolution runs, Ref-L025N0752 and Recal-L025N0752. Both runs use a box of  $(25 \text{ cMpc})^3$  and have better spatial and mass resolutions – by factors of 2 and 8, respectively – than the  $25 \text{ cMpc}^3$  intermediate-resolution run (Ref-L025N0376). Ref-L025N0752 uses the same subgrid parameter values as the Ref-L025N0376 run and is used for strong convergence tests, while Recal-L025N0752 is the re-calibrated run and is used for weak convergence tests.

We analyse the  $\text{H I}$  mass fraction of satellite galaxies as a function of their stellar mass  $M_{*}$  and host halo mass  $M_{200}$ . Given the smaller

volume, here we are limited to a narrower range of stellar and host halo masses and to a much smaller sample of galaxies with respect to the analysis done in Section 4. Therefore, we split our sample of satellites in four bins, determined by whether or not their stellar mass is greater than  $10^{9.5} M_{\odot}$  and by whether or not they reside in haloes more massive than  $10^{12.5} M_{\odot}$ . As before, we assume a minimum  $\text{H I}$  mass of  $1.36 \times 10^6 M_{\odot}$ , corresponding to the mass resolution of the Ref-L025N0376 run. In Fig. B1 we compare the *median*  $f_{\text{H I}}$  derived for the four bins in the runs Ref-L025N0376, Ref-L025N0752 and Recal-L025N0752.  $1\sigma$  error bars are derived by bootstrap resampling the galaxies in each bin. In all cases, the three runs show values of  $f_{\text{H I}}$  that are consistent with each other and predict a drop in the median  $f_{\text{H I}}$  with increasing  $M_{200}$ . This indicates that the mass resolution of the Ref-L025N0376 run – and therefore of the Ref-L100N1504 run – is adequate to model the physics of the environmental processes. In general, the runs at higher resolution predict slightly larger  $\text{H I}$  masses than the intermediate-resolution run, as also shown by Bahé et al. (2016) and by Crain et al. (2016).

This paper has been typeset from a  $\text{\LaTeX}$  file prepared by the author.